

PT encoded polypeptide, useful e.g. for regulating gene expression -
XX
XX Claim 11; Page 48-50; 74pp; English.
XX
CC The present sequence is the glutamyl-tRNA synthetase for corn. The enzyme
CC is an aminoacyl-tRNA synthetase (AARS). AARS charge (acylate) specific
CC tRNAs with amino acids for use in protein synthesis. Glutamyl-tRNA
CC synthetase therefore charges a glutamyl-tRNA with glutamate. Since this
CC enzyme has a crucial role in protein synthesis and therefore life, any
CC agent that inhibits or disrupts protein synthesis is likely to be toxic.
CC The present sequence could therefore be used as a basis for testing
CC whether the encoded aminoacyl-tRNA synthetase is sensitive to known
CC inhibitors or other chemicals and hence could be used in the discovery of
CC potential herbicides.
XX
XX Sequence 715 AA;

Query Match 100.0%; Score 3734; DB 21; Length 715;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 715; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEALSPSKDPSPIITICAKLVGLPLTINSLAAGSAPTOFASGESLHGVNPIITLIA 60
DB 1 MEALSPSKDPSPIITICAKLVGLPLTINSLAAGSAPTOFASGESLHGVNPIITLIA 60
QY 61 RGASIASISGKNDIEFGHVEWLEVAPTFISGSEFENACLFVDGFLASRTFLVGHGLTIA 120
DB 61 RGASIASISGKNDIEFGHVEWLEVAPTFISGSEFENACLFVDGFLASRTFLVGHGLTIA 120
QY 121 DIAVWSNAGIGQWESIRKSKYONLVWFNSIDSEYKALNEVVAFVKRGIGKSPA 180
DB 121 DIAVWSNAGIGQWESIRKSKYONLVWFNSIDSEYKALNEVVAFVKRGIGKSPA 180
QY 181 PSLKEKYVDSKDPAPAEVDLPQAKYKVCVFAPEPSGYLHIGAKALNKKYFAERYOG 240
DB 181 PSLKEKYVDSKDPAPAEVDLPQAKYKVCVFAPEPSGYLHIGAKALNKKYFAERYOG 240
QY 241 RLIVRPDDTNPSSKSNFEVENLKDIEFLGIKYDAVYTSOYFPKLMEMASLIIQOGKAY 300
DB 241 RLIVRPDDTNPSSKSNFEVENLKDIEFLGIKYDAVYTSOYFPKLMEMASLIIQOGKAY 300
QY 301 IDDPKKEKMRKRDGIESRCRNNTVEENLSLMEKNVNGTRGMOCCVRGKLDMDPKKS 360
DB 301 IDDPKKEKMRKRDGIESRCRNNTVEENLSLMEKNVNGTRGMOCCVRGKLDMDPKKS 360
QY 361 LRDPVYVRCNTDPHHRVGSKYKVPYDFACPFVDALSGVTHALRSSSEYHNRNAQYRIL 420
DB 361 LRDPVYVRCNTDPHHRVGSKYKVPYDFACPFVDALSGVTHALRSSSEYHNRNAQYRIL 420
QY 421 QDMGLRVEIYEFSLNNVYTLSSRKLLMFPVONKKVEDWTDPRPTVOGIVRGLKVEA 480
DB 421 QDMGLRVEIYEFSLNNVYTLSSRKLLMFPVONKKVEDWTDPRPTVOGIVRGLKVEA 480
QY 481 LIQFLIOGASRNUNLMEWDLMTINKKIIDPVCARHRAVLKDOQVITTLNPGPEBPVR 540
DB 481 LIQFLIOGASRNUNLMEWDLMTINKKIIDPVCARHRAVLKDOQVITTLNPGPEBPVR 540
QY 541 ILPRHKEGEGAKATTFANRIMLDYADAAAINKEEETLMDGNAIYKEIKVESGVTTE 600
DB 541 ILPRHKEGEGAKATTFANRIMLDYADAAAINKEEETLMDGNAIYKEIKVESGVTTE 600
QY 601 LVGSLHLEGSVKTTLKLTITWLADIBELVPLSVEFDYLSIKKKLEDEDFLDNLNPTCRR 660
DB 601 LVGSLHLEGSVKTTLKLTITWLADIBELVPLSVEFDYLSIKKKLEDEDFLDNLNPTCRR 660
QY 661 EIPALGDANMNRIKGEIITOLERKGYRCDAFITSSRPVULFAIPDRQOASLS 715
DB 661 EIPALGDANMNRIKGEIITOLERKGYRCDAFITSSRPVULFAIPDRQOASLS 715

RESULT 2
AAG41411
ID AAG41411 standard; Protein; 716 AA.

XX AC AAG41411;
XX 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 51519.
XX KM Protein identification; signal transduction pathway; metabolic pathway;
XX KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX OS Arabidopsis thaliana.
XX PN EPI033405-A2.
XX PD 06-SEP-2000.
XX PE 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123160.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.
XX PR 28-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 04-MAY-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
XX PR 10-JUN-1999; 99US-0138847.
XX PR 14-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139492.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139455.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.
XX PR 18-JUN-1999; 99US-0139459.
XX PR 18-JUN-1999; 99US-0139460.
XX PR 18-JUN-1999; 99US-0139461.

QY	229	Q6RLVRFEDDTPTPSKSNSEFVENLLADIETLGIKYDAVVTYSYFPLKMAASLIYKQK	236
Db	242	Q6EIVRFEDDTPTPAKSNSEFVDLVKDIGTLGIKYETVITYSYFPELMAAEKLMMEGK	301
QY	249	AYIDTTPKQMKERKRDGIESRCRNNTVBEENLSLMKEMVNGTERGMQCCVCRGLKMDQPN	358
Db	302	AYVDTPRQGMQKERDGIIDSKCRNHSVEENLTKLMEMINGSERGLDCCRGKFPNQDPN	361
QY	359	KSRLDPVYVRCNTDPHHRVSGKKYVPTYDFACPFDALLEGVTHALRSSEYHDMQAQYR	418
Db	362	KAMRDPVYVRCNPMSHRIGDKYKIYPTVDFACPFDVDSLEGITHALRSSEYHDMQAQYF	421
QY	419	ILQDQGLRVEIYEFSRLMMVYTLKSKRLKLMFVNQKKVEBDMTPDPRPTVQGIYRGLKY	478
Db	422	VLEDQGLRVQLYEFSRLNDVFTLLSKRKLKLMFVQGLVGMQDPRPTVQGIYRGLKI	481
QY	479	EALIQFIILQOGASKNLTLMEMWDLKMTINKKIIPVCAHRTAVLKDQVIFTLTNGPEEP	538
Db	482	EALIQFIILQOGASKNLTLMEMWDLKMSINKKIIPVCPRHRTAVVAERVLFTLTDGPEEP	541
QY	539	VHILPRHKPEEGAGKATTPANRIMLDYADAAINKGEVYTLMDWGNAIYKEI-KVESGV	597
Db	542	VMIPIPHKKPEEGAGKATTPFTKSIWLEADASAISVGEETVLTMDWGNAIYKEITKDEGR	601
QY	598	IREIVGELHLESGVKTITKLKITWLADIEELVPLISVFDYLLSKKLEEDDFLNDLNP	655
Db	602	VALLSGVNLVQSGVKTITKLKITWLPTNELVNLITLEFDYLLTKKLEEDDEADVNP	661
QY	658	TRREIIPALGDANNRNIKGEIILQERKGYRCDAPIFRSSKPVVLFPAIPDGR	709
Db	662	TKKETLALDSDNMRNLKCGDVILERKGYRCVPIFKSSKPIVLFSPDGR	713
RESULT 3			
ID	AAG41410	standard; Protein; 719 AA.	
XX	AC	AAG41410;	
XX	DT	18-OCT-2000 (first entry)	
XX	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 51518.	
KM	XX	Protein identification: signal transduction pathway; metabolic pathway;	
KM	XX	hybridization assay; genetic mapping; gene expression control; promoter;	
OS	XX	termination sequence.	
OS	XX	Arabidopsis thaliana.	
PN	XX	EP1033405-A2.	
XX	PD	06-SEP-2000.	
XX	PF	25-FEB-2000; 2000EP-0301439.	
XX	PR	25-FEB-1999; 99US-0121825.	
XX	PR	05-MAR-1999; 99US-0123180.	
XX	PR	09-MAR-1999; 99US-0123548.	
XX	PR	23-MAR-1999; 99US-0125788.	
XX	PR	25-MAR-1999; 99US-0126264.	
XX	PR	29-MAR-1999; 99US-0126785.	
XX	PR	01-APR-1999; 99US-0127462.	
XX	PR	06-APR-1999; 99US-0128334.	
XX	PR	08-APR-1999; 99US-0128714.	
XX	PR	16-APR-1999; 99US-0129845.	
XX	PR	19-APR-1999; 99US-0130077.	
XX	PR	21-APR-1999; 99US-0130449.	
XX	PR	23-APR-1999; 99US-0130510.	
XX	PR	23-APR-1999; 99US-0130691.	
XX	PR	28-APR-1999; 99US-0131449.	
XX	PR	30-APR-1999; 99US-0132048.	
XX	PR	30-APR-1999; 99US-0132407.	

PR	05-MAY-1999;	9905-0133484;
PR	06-MAY-1999;	9905-0133486;
PR	06-MAY-1999;	9905-0133487;
PR	07-MAY-1999;	9905-0133488;
PR	11-MAY-1999;	9905-0134256;
PR	14-MAY-1999;	9905-0134218;
PR	14-MAY-1999;	9905-0134221;
PR	14-MAY-1999;	9905-0134370;
PR	18-MAY-1999;	9905-0134768;
PR	19-MAY-1999;	9905-0134941;
PR	20-MAY-1999;	9905-0135124;
PR	21-MAY-1999;	9905-0135353;
PR	24-MAY-1999;	9905-0135629;
PR	25-MAY-1999;	9905-0136021;
PR	27-MAY-1999;	9905-0136392;
PR	28-MAY-1999;	9905-0136782;
PR	01-JUN-1999;	9905-0138504;
PR	03-JUN-1999;	9905-0137528;
PR	04-JUN-1999;	9905-0137524;
PR	07-JUN-1999;	9905-0137702;
PR	08-JUN-1999;	9905-0138094;
PR	10-JUN-1999;	9905-0138540;
PR	14-JUN-1999;	9905-0138647;
PR	16-JUN-1999;	9905-0138119;
PR	16-JUN-1999;	9905-0133452;
PR	16-JUN-1999;	9905-0133453;
PR	17-JUN-1999;	9905-0133452;
PR	18-JUN-1999;	9905-0133454;
PR	18-JUN-1999;	9905-0133455;
PR	18-JUN-1999;	9905-0133456;
PR	18-JUN-1999;	9905-0133461;
PR	18-JUN-1999;	9905-0133463;
PR	18-JUN-1999;	9905-0133462;
PR	18-JUN-1999;	9905-0133750;
PR	23-JUN-1999;	9905-0140695;
PR	24-JUN-1999;	9905-0140921;
PR	28-JUN-1999;	9905-0141287;
PR	30-JUN-1999;	9905-0141854;
PR	01-JUL-1999;	9905-0142154;
PR	02-JUL-1999;	9905-0142055;
PR	06-JUL-1999;	9905-0142390;
PR	08-JUL-1999;	9905-0142803;
PR	09-JUL-1999;	9905-0143920;
PR	12-JUL-1999;	9905-0142977;
PR	13-JUL-1999;	9905-0143544;
PR	14-JUL-1999;	9905-0143624;
PR	15-JUL-1999;	9905-0144005;
PR	16-JUL-1999;	9905-0144086;
PR	19-JUL-1999;	9905-0144325;
PR	19-JUL-1999;	9905-0144331;
PR	19-JUL-1999;	9905-0144332;
PR	19-JUL-1999;	9905-0144333;
PR	19-JUL-1999;	9905-0144334;
PR	19-JUL-1999;	9905-0144335;
PR	20-JUL-1999;	9905-0144352;
PR	20-JUL-1999;	9905-0144632;
PR	21-JUL-1999;	9905-0144814;
PR	21-JUL-1999;	9905-0145086;
PR	21-JUL-1999;	9905-0145088;

PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 71.2%; Score 2657; DB 21; Length 719;
Best Local Similarity 69.0%; Pred. No. 1,66-226;
Matches 491; Conservative 106; Mismatches 107; Indels 8; Gaps 5;

QY 5 LSFESKSPPIITCAALVGLPLTINSLAAGAPTLQFASGESLHGVNPIITYARCA- 63
DB 6 LSFPESEPLSLVIALSLASAPVITDSSAAATVPSEFVFSGRKXNGATVLLRYVGRSAK 65
QY 64 SIASLSGKNDIEFPHVWEMLEAPTLPLSGSEFFENACLVFDGFLASRFTLVGHGLTIADIA 123
DB 66 KLDPFYGNNAFDSQIDWVDYASVFSSGSFFENACGRVDYALSSFTLVGHSLSIADIVA 125
QY 124 VMSNLAGIGQRWESLRSKSKYQNLVRFNSIDSEYKEALNEVNAFYGKRIGRS-PAPS 182
DB 126 IMSALAGTQGRWESLRSKSKYQNLVRFNSIDSEYKEALNEVNAFYGKRIGRS-PAPS 184
QY 183 LKEKVH---DSKPSAPEVDLPEAKVGVKCVRPAPESGYLHIGHAAALLNKYPARY 238
DB 185 SKDSQAVKGDQKGRPEVDLPEALIGKVLRFAPESGYLHIGHAAALLNKYPARY 244
QY 239 QGRLLVFPDDTPNPKESNEFENLKDIENTGICYDAVTYSDYEPKIMEAESLKGCK 298
DB 245 QGEVIVRPDDTPNPKESNEFENLKDIENTGICYDAVTYSDYEPKIMEAESLKGCK 304
QY 299 AYIDTPKEQMRKRMGIESRCRNNTVEENLSLMKEVNGTERGMOCCVYRGLDMODPN 358
DB 305 AYVDTPREQMRKRMGIDSKRNHSEVENLKLKMEHIASERGLQCCVYRGMODPN 364
QY 359 KSLRDPVYRRCNTDPHHRVSGKYVPYTPDFACPFVDALSGVTHALRSSEYHDAQAQYR 418
DB 365 KAMRDPVYRRCNPMHSHRIGDKYKILPYDFACPFVDSLEGITHALRSSEYHDAQAQYR 424
QY 419 ILDMGLRVEIYFESRLNMYTTLISKRKLIMFVQNKKEVDWTPRPFTVGGIYRGLKV 478
DB 425 VLEDMGLRVOVLFEFSRLNMYTTLISKRKLIMFVQNKKEVDWTPRPFTVGGIYRGLKI 484
QY 479 BALIOTILQOGASKNLIMEDKLTMTINKKIIDVCAKHTAVLKDORVIFLTLNGPEEPF 538
DB 485 BALIOTILQOGASKNLIMEDKLTMTINKKIIDVCAKHTAVLKDORVIFLTLNGPEEPF 544
QY 539 VRIILPRHKKFEGAGKKAATTFANRIMLDYADAAAIKGBEVLTMQNAIVKEI-KVESGV 597
DB 545 VRIILPRHKKFEGAGKKAATTFANRIMLDYADAAAIKGBEVLTMQNAIVKEI-KVESGV 604
QY 598 ITBLVGBLHEGSKYTKTKLTLADIBELVPLSLVEFDYIISKKKLEBDEDFLDNLNCP 657
DB 605 VTALSGVNLQGSYKTKTKLTLADIBELVPLSLVEFDYIISKKKLEBDEDFLDNLNCP 664
QY 658 TRREIPALGANMNRINRGEIOLERKGYVCDAPFIRSSKPYVLPALIPDGR 709
DB 665 TKKETTLALGDSNMNRINRGEIOLERKGYVCDAPFIRSSKPYVLPALIPDGR 716
RESULT 4
AAG41409

ID AAC41409 standard; Protein; 748 AA.
XX AAC41409;
AC
XX 18-OCT-2000 (first entry)
DT
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51517.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX Arabidopsis thaliana.
XX EPI033405-A2.
PN
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138549.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147392.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149358.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.

```
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0158295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 71.2%; Score 2657; DB 21; Length 748;

Best Local Similarity 69.0%; Pred. No. 1.7e-226; Mismatches 107; Indels 8; Gaps 5;

Matches 491; Conservative 106; Mismatches 107; Indels 8; Gaps 5;

```
OY 5 LFSKDSPIRISIIICAKLVGLPLTINHSIAGSAPTLQFASGESIAGVNPITIIYARGA- 63
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 35 LSPFSPSPPLVIVALSISASPVITIDSSAAATTVSPFSPDRKLNKATVILIRYGRSAK 94
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 64 SIASISGKNIDIEFGHVEWLEVAPTFLSGSEFENACLFVDGELASRTFLVGHGLTIADIA 123
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 95 KLDPFYGNNAFDSQIDIDEMVDVYASVSSGSEFENACGRVDKYLESTFLVGHSLIADVA 154
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 124 VMSNLAGIGORMWESIRKSKKONTVRWFNSIDSEKELNEVVAFVGKRGIGKS-PAKS 182
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 155 IWSALAGIGORMWESIRKSKKQSLVRWFNSIIDSEYSEVINKLATYV-KKSGKRVAAK 213
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 183 LKEKVH----DSKSDASAEVDLPGAKVGKVCVRPAPEPSGYLIHGAKAALLNKYFABRY 238
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 214 SKDSQAAVKGDDQDKGKEVNDLPEAEIGKVKLRAPAPBSGYLIHGAKAALLNKYFABRY 273
OY 239 QGRILVFPDDTNPESKESNEFENLLKDIETTGIXYADATYSDYFKLMEAESLIIKGGK 298
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 274 QGEVIVAFDDTNPAKESNEFVDNLVKDITGIIYKTYTSDYFPPELMDPAEKLMRGK 333
OY 299 AYIDDTPEQMRKEMDIESRCNNYVENLISLMEKVNTERGMQCCVARGKLDMDOPN 358
Db 334 AYVDOTREQMQKERMGDIDSKRNHSEVENLKLMEKMIAGSEGLQCCVARGKFMQDPN 393
OY 359 KSLRDPYTYRCNTDPHHRVSGKIVYPTYPACFPVDALBSVTALNSSEYHDNRQAQYR 418
Db 394 KAMDPPVYYRCNPMSHRIGDKYKIPYDPAFCFVDSLBSGITHALNSSEYHDNRQAQYF 453
OY 419 ILDMGLRVEIYFESRLNMYTLLSKRKLMPFONKKBWPTPRPFTVGGIYRGGLKY 478
Db 454 VLEDMLRQVOLYFESRLNMYTLLSKRKLMPFQGLVDMDDPRPFTVGGIYRGGLKI 513
OY 479 EALLIQTLOQASQNLNLMEMDKLMTINKIIDPVCARHTAVLKDORVIFLTNGPEBEP 538
Db 514 EALLIQTLOQASQNLNLMEMDKLMSINKRIIDPVCARHTAVLKDORVIFLTNGPEBEP 573
OY 539 VRIIPRHKKFEAGKAKATTANRIWLDYDAAAIKGBEVTLMQNAIYKEI-KVESGV 597
Db 574 VRMI PKHKKFEAGKAKATTFTKSIWLEBADAISVGEVTLMDMGNAIYKEITKDEBGR 633
OY 598 ITTELVEGLHLEGSVKTTRKLTWLAJIBELVPLSLVPEVDYLSKKKLEDEDFLDNINPC 657
Db 634 VTALSGVNLQGSVKTTRKLTWLAJIBELVPLSLVPEVDYLSKKKLEDEDFLDNINPC 693
OY 658 TRREIPALGDAMNENIKRGEIIQLEKRGYRCDAFIRSSKPVVLAIPDGR 709
Db 694 TKKETLALGDSNMNMLKCGVDIQLERKCYFPCDVPFKVSSRPYLFSLPDGR 745

RESULT 5
ABP73689
ID ABP73689 standard; Protein, 725 AA.
AC ABP73689;
XX 30-JAN-2003 (first entry)
XX
XX Candida albicans essential protein SBQ ID NO 7526.
XX
XX Fungus; yeast; tetracyclin, promoter; GRACB strain; bioethnthesis;
XX signal transduction; DNA replication; cell division; growth;
XX proliferation; Candida albicans; fungicide; antifungal.
XX
XX Candida albicans.
XX
XX WO200253728-A2.
XX
XX 11-JUL-2002.
XX
XX 26-DEC-2001; 2001WO-US49486.
XX
XX 29-DEC-2000; 2000US-259128P.
XX 20-FEB-2001; 2001US-0792024.
XX 22-AUG-2001; 2001US-314050P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL,
XX WPI; 2002-56694/60.
XX N-PSDB; AB232239.
XX
XX Constructing strains for identifying gene products as effective targets
XX for therapeutic intervention, by inactivating in the strain one allele
XX of a gene and placing other allele of the gene under conditional
XX expression -
```


QY	177	KSPPASLKE---KVNDSKDPSPAEVLLPQAKVQKVCVRAPRPSGVLHGHAAALANK	223
Db	171	RSPOSSKEQTPAKTGRKQ-EGKFVLLPGAEWKGVVVRPFPASGYLHGHAAALINQY	229
QY	234	FAERVOGRLLVFPDDTNPSKESNEFEVNLKDIETLGIKYDAVTYSDVPEPKMEMAESL	293
Db	230	YALAQVQTLIMFPDDTNPAKETEVEFENVLIGDEQIQIKPDVYTHSNFDMLDYCVRL	289
QY	294	IKQGAQYIDITPKEQNRKERMDGISRCRNNTVEENLSLWKEMVNGSTERGMOCCVRAKLD	353
Db	290	IKESKAYVDDTPPEQMKLEREQVESANSNSVERNKLSLWEEVVKSGEKQKVCVRAKID	349
QY	354	MODPKNSLRDPVYYRNTDPNHRVSGKTKYKVPYDPAACFPVDALLEGYTHALASSEYHDN	413
Db	350	MSSPFGCMMDPTIYRCNBPHPRTGKTKYKVPYDPAACPIVDIAIENVYHTLRTEYHDDR	409
QY	414	AQYRLLQDMGRLRVAYIEFESRLNMYTLLSKRKLMPVONKQVEDMTDPRPTVOGIYR	473
Db	410	DOFYFPIIDLKIRKPIWYSYSLNMTNTVLSKRKLTFWFDOSGLVDGMDDRPFTVGIIR	469
QY	474	RGLKYEALIQTLQOGASKNLMLMEWDKLTNTNKKLIIDPVCAHTAVLKDORYIFTLTNG	533
Db	470	RGMTVEGLKEFPIIAQSSSKSVFPMWMDKLIWAFKAYITDIPARPYTLBLEKERYIVANVAGA	529
QY	534	PEEPFVRI-LPRHKKFEGAGKCAATTANRIWLDYADAAALNKGEVTLMDWGAIVKEI-	591
Db	530	KVE---RIGVSVHPKDESIGKKTVLLGPRIYIDYVDAELKEGENATPTINWGILIRKN	586
QY	592	KVESGVITELVGEIHLDEGSVKTKTKIITMLA---DIEELVPLSLVEFDYLSKKLEED	648
Db	587	KDASGNITSVDAALNLENKDFKKTLLTMLAAVEDDSAYPPFTFCVYFDNITSKAVLGKB	646
QY	649	DFLDNLNPTCRREIPALGDANMNRNIKRGEIIQLERKGYRCDAPIRRS-----KPVLL	702
Db	647	DFKQPIGHTTIREVPLGDPBELKCKCKGDILOLGRGFGFRVDAVYAPPSGYTNVPSIYL	706
QY	703	FAIPDG 708	
Db	707	FSIPDG 712	
RESULT 7			
AAU04349	ID	AAU04349 standard; Protein; 1512 AA.	
XX	AC	AAU04349;	
XX	XX	23-OCT-2001 (first entry)	
XX	XX	Mammalian toxicological response marker protein #1.	
XX	XX	Mammalian toxicological response marker; antigen; antibody; agonist.	
XX	XX	Homo sapiens.	
XX	XX	WO200136684-A2.	
XX	XX	25-MAY-2001.	
XX	XX	16-NOV-2000; 2000MO-US31743.	
XX	XX	19-NOV-1999; 99US-0443184.	
XX	XX	(INCY-) INCYTE GENOMICS INC.	
XX	XX	Cunningham MJ, Zweigler GB, Kaser MR, Panzer SR, Seilhammer JI;	
XX	XX	Yue H, Baughn MR, Azimzai Y, Lal P;	
XX	XX	WPI; 2001-355646/37.	
PT	PT	Novel mammalian nucleic acid molecules whose levels are up regulated or	
PT	PT	down regulated following treatment with a toxic compound, useful for	
PT	PT	detecting metabolic and toxicological responses and in monitoring drug	

PT action -
XX
PS Claim 9; page 56-59; 89pp; English.
XX

The sequence represents a novel mammalian protein encoded by a nucleic acid molecule whose levels are up regulated or down regulated following treatment with a toxic compound. Polynucleotide sequences complementary to the sequences of the invention are useful for preventing a toxicological response by acting against one or more up-regulated nucleic acid molecules. An agonist identified by the above method is also useful for preventing a toxicological response by initiating transcription of a gene comprising a down regulated nucleic acid molecule of the invention. Proteins encoded by the nucleic acids of the invention are useful for producing an animal model system. The nucleic acid molecules are useful for producing an animal model system. The nucleic acid microarray format may be used to characterize gene expression patterns associated with novel compounds to elucidate any toxicological responses, or to monitor the effects of treatments during clinical trials or therapy where metabolic response to toxic compounds may be expected.

The nucleic acids are useful for various hybridisation technologies and is useful for designing hybridisation probes. The nucleic acid molecule or its fragment, or a protein encoded by the nucleic acid molecule may be used to purify a ligand from a sample.

SQ Sequence 1512 AA;

Query Match 40.3%; Score 1506; DB 22; Length 1512;
Best Local Similarity 42.2%; Pred. No. 8,2e-124;
Matches 305; Conservative 146; Mismatches 232; Indels 40; Gaps 9

4 ALSFSDSPPIISIIICAAKLVGLPLTINSHLSAASAPTLQFASGESLHGVPYIITYARGA 63
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
5 SLTVNSGPRPGALLAVHENVDDSV--SVEEGKNILHVSENVFTDYNSTLRYLARA 62
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
64 STATSGKNDIEFGHVWLEAYPFLSGSFENACLFVDGFLASRTVLVGHGLTIADI 122
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
63 TTAGLYGSNLMHEHTIDHWLEFSATKLSCDSFSTITNELNLCLSLRTVLVGNSTLSLDL 122
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
123 AVMSNLAIIGRWESLRSKKYQLNVRWNNSIDSEYKELNVAFAVGKRGKGPSAPS 182
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
123 CWAVALTKNAAMQEQLOKKRAVHVKRWFGELEAO-QAFQGV-----G 164
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
183 LKEATHSDKSAPR-----DLPAKVGVKVCFRFPSPGYLIHGAKAALINKFY 234
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
165 TKWDSTTKARVAREPKODVGFYBLPGAEMKVVRPREPASGYLIHGAADALLNGHY 224
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
235 AERVOGRLLIVRPDTPSKESNEPVENLIKDIETLGIKYDAVTYSDFPKLMEWAESLI 294
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
225 QVNFKGLIMRFDDTNPEKEDEFKVLIEDVAMHLIKPDQPTYSNHDEFTIMKYAEKLI 284
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
235 KQGKYYIDDTPREQWRKERMGDISERCNNNYEENLSLMKEWVNGFERMOCVCVRGKLD 354
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
285 QEGKYYVDTPAEQMKAREORIBESKHKNPLEKNLQWMEBKKSQFQSCCLRAKIDM 344
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
355 QDPNKSLDPPVYYCNPDPHHRVGSKYKYPDYDACEFVDLAEGTYHALRSSEYNDRNA 414
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
345 SSNNCGMRDPLLRYRKIKIPHRITGKNKVVYPTYPDACPVSIBGTTHALKRTTEHNDRB 404
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
415 QYRIILODMGLRRVEIYEFSRLNMVYTLLSRKKLLMFYONKKVLEDWTDERFFPTVGIVAR 474
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
405 QFYWIIEALGIRKPFIWEYSRLNLTNNVLSKRKLTFWVNEGJLVGDMDPRPFYVGVLRR 464
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
415 GLKYEALLQFLIOGASAKNMLMEMDKLMTINKLIIDRPCAHNTAVLKQRVIFPLTNGP 534
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
465 GMTVEGLKQFIALAQSSSRSSVANMEMDKIWAFFKVIIDPAPPYVALLKKEVLPANVPEAQ 524
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
535 EEPFRILIPRHKKFEAGAKKATTFNRIWLVDADAALINKGEVYILMDGNALIVEI-KV 593
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
525 EE-MKEVAKHPKPEVGLKPYWPSPKPIFBADADETSBGEMVTFINMGNIINTIKIHON 582
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
594 ESGVITELVGEHLHGGSVKTKTKITWLADIEELVPLSL--VERDYLLSKKLLEDEDFL 651
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::

Db 583 ADGKIISLDAKLNLKNDYKTKTKVTWLAETHALPIVICVYEHILITKPVLGKDEDFK 642
QY 652 DNLNPTTRREIPALGDANMRNIRKGEIITOLERKGYRCD-----APFIRSSKPVVLPFAI 705
Db 643 QYVKNNSGHEBLMLGDPLCKDKLKKGDIIOLORRGFICDPYEPVSPYSCKEAPCVLIYI 702
QY 706 PDG 708
Db 703 PDG 705

RESULT 8
AAM78732 standard; Protein, 1512 AA.
XX AAM78732;
XX AAM78732;
XX 06-NOV-2001 (first entry)
XX Human protein SEQ ID NO 1394.
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX Homo sapiens.
XX MO200157190-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US04098.
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 30-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX Tang YF, Liu C, Dimauc RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y,
PI Zhao QH, Wang D, Zhang J, Ren F, Chen R, Wang ZW,
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R,
XX WPI; 2001-476283/51.
XX N-PSDB; AAK51865.
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX Claim 20; Page 3654-3657; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAW80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX Sequence 1512 AA;

Query Match 40.2%; Score 1502; DB 22; Length 1512;
Best Local Similarity 42.0%; Pred. No. 1,96-123;
Matches 304; Conservative 147; Mismatches 232; Indels 40; Gaps 9;
QY 4 ALSPKDSPPISIIICAKLVGLPLTINHSLSAAGAPLQFASGSLGVNPIIYIARGA 63
Db 5 SLIVNSGDPFLGALLAVHVDVSI--SVLEGKENTLHSENVIFPDVNSILNKLARVA 62
QY 64 SIASLGNKDIIEFGVVEWLEAPFLSGSE--FENACLFDVDFLASRTFLVGHGLTADI 122
Db 63 TTAGLYGSNLMHEHTIEHMLEFSATKLSGDSFSTINEINLHCILSTRYLVGNSLSADL 122
QY 123 AVMSNLAGIGORMSLSKSKYQNLVWVENSIDSEYKALNEVYAAFPVGRGIGKSPRS 182
Db 123 CVMATLKGNAAMQOLKQKAPVHVKKWFFLEAQ--QAFQSV-----G 164
QY 183 LKEKYHDSKDSAPE-----VDLPQAKGVKCVFAPFAPSPGYLHGHAKAALLNRYF 234
Db 165 TKMDVSTTKARVAPEKKQDVGFELFGAEMGKTVAFPPFASGYLHGHAKAALLNRY 224
QY 235 AERYQGLIVAFDDTNPSSKESNEFVENLADIEITLGIKYDAVYTSQVFPPLMEMASLI 294
Db 225 QVNFKGLIMRFDDTNEKKEDFEKYLIEDVAMLHIKPDQFTYSDHFEITMKYAEKLI 284
QY 295 KQGAIVDDTPKEQKREKRDGIESRCRANTVEENLSLMKEMNANGTERGMQCCVGRGLDM 354
Db 285 QEGRAVYDDTPAEQKREBQRISSKRRKPIEKNLQWMEEMKKSQFGSGCLRAKIDM 344
QY 355 QDPKSLRDPVYVRCNTDPHHRVGSKYKVPYFPACFPVDALRGVTHALSSEYHDBNA 414
Db 345 SSNNGCRDPTLYCKIOPHRTGKKNVPTYPACPIYDSIGCVTHALRTTEYHDBE 404
QY 415 QYRILQDMGLRVEIYEFSLNMYTLLSKRKLMPVQNKVEDWTDPRFPYQGIIVR 474
Db 405 QFYWIEALGIRKRYIWEYSRLNNTVLSRKLTWTFNEGLVGMDDPRFPYRGVLR 464
QY 475 GLKYPALIOFTLQOGASKNLMLMDKMTINKKIIDPVCARHAYLKDQGVIFTLTNGP 534
Db 465 GMTVEGLKQFTIAOGSSRSVVMEMDKIMENKVIIDPVARVYALLKKEVIPPVNVBAQ 524
QY 535 EEPFVRLPRHKKEGAGKATTEFANRIMLDYADAALINKSEETLMDMGNAIYKEI-KV 593
Db 525 EE--MKSVAKHPKPEVGLKPVWYSPKVFIEGADAEFTSESEMTFIMGNLNTTKIKN 582
QY 594 ESGVITBELVGLHLEGSVTKTKLKITWLAIDIELVPLSL--VEBDYLSKKLSEDEDFL 651
Db 583 ADGKIISLDAKLNLKNDYKTKTKVTWLAETHALPIVICVYEHILITKPVLGKDEDFK 642
QY 652 DNLNPTTRREIPALGDANMRNIRKGEIITOLERKGYRCD-----APFIRSSKPVVLPFAI 705
Db 643 QYVKNNSGHEBLMLGDPLCKDKLKKGDIIOLORRGFICDPYEPVSPYSCKEAPCVLIYI 702
QY 706 PDG 708
Db 703 PDG 705

RESULT 9
AAM79716 standard; Protein, 1550 AA.
XX AAM79716;
XX 06-NOV-2001 (first entry)
XX Human protein SEQ ID NO 3362.
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX Sequence 1512 AA;

OS Homo sapiens.
XX WO200157190-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US04098.
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX Tang YF, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y,
PI Zhao Q, Wang D, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX WPI; 2001-476283/51.
XX N-PSDB; AAK52849.
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX Claim 20; Page 312; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoietic regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAW80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX Sequence 1550 AA;
XX
XX Query Match 40.0%; Score 1493; DB 22; Length 1550;
XX Best Local Similarity 41.9%; Pred. No. 1,2e-122;
XX Matches 303; Conservative 147; Mismatches 233; Indels 40; Gaps 9;
QY 4 A L S F S K D S P P I S I C A K L V G L P L T I N H S L A G S A P T L Q F A S G E S L H G V N P I I L Y I A R G A 63
DB 43 S L T V N S G D P P I G A L L A V E H V K D V S I - S V E G K E N I I L H S E N V I F T V N S I R L A V A 100
QY 64 S I A S I S G K N D I E F G H V W E L E A P P T L S G S E - F E N A C I F V D G F L A S R F L V G H G I T I A D I 122
DB 101 T T A G I Y G S N L M E H E I D I M L E F S A T K S C D S F T S T I N E L N H C I S L R T Y L V G N S I L S L D L 160
QY 123 A V M S N L A G I G R W E S L R K S K K Y O N L V R W F N S I D S E Y K A L M E V N A F A F G K G I G S P A S 182
DB 161 C W A T L K N A M A Q E D L K O K A P A V H K M F G L E A Q - Q A F Q S V - - - - - G 202
QY 183 L K E R V H D S K D S A P E - - - - - V D L P G A K V G K V C R P A P E P S G Y L A I G H A K A A L L N K Y F 234
DB 203 T K M D V S T T K A R A P E K O D V G F V E L P G A E M G K V T V R P P E A S G Y L H I G H A K A A L I N G H Y 262
QY 235 A E R V O G R I V A P F D D T N P S K E S N E F E N I L K D I E T I G I T A D V T T Y S D Y P P K L M E A S E L I 294
DB 263 Q V N F G K L I M F P D D T N P E K E D E F E K V I L E D V A M L H I P R D P T Y T S D H F E T I M K Y A E K L I 322
QY 295 K O G K A Y I D D T P R E Q R K E R M D G I E S R C A N N T V E E N L S L M K E M V N G T E R G M Q C V R G K L D M 354

DB 323 Q E G R A Y V D D T P G E O I K A R E O R I E S K H R K N I E K N L Q W E E K K G S Q F G H S C C L F A K I D M 382
QY 355 Q D P K S L B D P V Y R C N D P H H R V G S K Y V P T Y P A C P F V D A L B G V T A L S S E V H D N A 414
DB 383 S N N G C M D P T L Y C K I Q P H P R T G K Y N V P T Y P A C I V D S I E G V T A L T T E Y H R D E 442
QY 415 Q V Y R I L Q D M G L R V E I Y E F S R L N M V Y T L S K R K L M F V O N K V E D M T P R P P V O G I V R 474
DB 443 Q F V I I E A L G I R K Y I M Y S R L N I N T V L S R K L T M F N E B L V D C M D P R P T Y R G V L R 502
QY 475 G L K V E A L I Q F L L O Q A S K N L M E M D K I T I N K K I I D P V C A R H T A V L K D Q R V I F T L T N G P 534
DB 503 G M T V E G L K O F I A O G S S R S V N M E M D K I W A N K K V I D P V A D R V A L L K E V I P A V N V P A Q 562
QY 535 E E P F R I L P R K K E F G A K K A T T P A N R I W D Y A D A A A I N K E E Y L M W G N A I Y E I - K V 593
DB 563 E E - M K E V A K H P K D E V G L K P W Y S P K V I G A D A E T F S E G E M V T F I M W G I N I T K I H K N 620
QY 594 E S G V I T E L V G E L H E G S V K T T K L I T M L A D I E E L V P L S L - V E P D Y L S K K K L E E D E D P L 651
DB 621 A D G K I S I D A K L N I E N D Y K K T T K Y T M L A E T T H A L P I V I C V T Y E H L T T K E Y L G K D E D P K 680
QY 652 D N L N P C T R R E I P A L G A N M R N I K G E I I Q L E R K G Y R C D - - - - - A P I R S K P V L P A I 705
DB 681 Q V N K N S K H E E L M G D P C L K D L K K G D I I Q L G R G F P I C D Q Y E P A V S P Y S C K E A P C V L A Y I 740
QY 706 P D G 708
DB 741 P D G 743
RESULT 10
ABU25717
ID ABU25717 standard; Protein; 715 AA.
XX ABU25717;
AC 16-APR-2003 (first entry)
XX DE Aspergillus fumigatus essential gene protein #375.
XX KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
XX KW cancer; contamination; biofilm; antibody; immune response.
XX OS Aspergillus fumigatus.
XX WO200286090-A2.
XX 31-OCT-2002.
XX 23-APR-2002; 2002WO-US13142.
XX 23-APR-2001; 2001US-285697P.
XX 27-APR-2001; 2001US-287066P.
XX 05-JUN-2001; 2001US-295890P.
XX 09-JUL-2001; 2001US-303899P.
XX 31-AUG-2001; 2001US-316362P.
XX (ELIT-) ELITRA PHARM INC.
XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX WPI; 2003-093124/08.
XX New purified or isolated nucleic acids of essential genes of
XX Aspergillus fumigatus, useful for treating or preventing infections by
XX A. fumigatus, or for treating a non-infectious disease in a subject
XX e.g. cancer -
XX Disclosure; Page -; 175pp; English.
XX The invention relates to novel purified or isolated nucleic acids of

CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organisms invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This sequence represents a protein of one of the essential genes
 CC of *Aspergillus fumigatus* of the invention.

XX Sequence 715 AA;

Query Match 39.2%; Score 1465.5; DB 24; Length 715;
 Best Local Similarity 42.9%; Pred. No. 9.9e-121; Indels 49; Gaps 14;
 Matches 309; Conservative 128; Mismatches 235;

4 ALSEKSPSPISIIICAKLVGLPTINHSLAAGAPLQF--ASGSEHGVNPILLYAR 61
 26 SINEKSPSPVISTVEDTAV-----LREGDAVAVQYAGASNPFGILINAVQELRK 76
 62 GASIASLSGRND-IFGHVVEWLEVAPTF--LSGSEFENACIFVDGFLASRTFLVGHQT 118
 77 --DEPFLNSKDEKE---NEMLSQLEAFAPLDFKALDPELQRLDTHLLRSFVVGYALS 130
 119 IADIIVNLSLAGIGQRMESLRKSKYQNLVWFNSID-----SEYKALNEVVAAPYK 172
 131 TADIALKCAIRGNRYAVALIKKG-SLVAVTRMFTFLBDLCPMAISTLEVLQAVR----- 184
 173 RGIGKSPAPSLKEKVDKSPAPAEVDLPGAKVGVCFAPAPSPGYLHIGAKAALINK 232
 185 -----EKKAKEGASGYDIALNTEKG-VVTRFPPEPGYHIGHAKAALINK 232
 233 YFA-ERYQRLIVRFDTPNPSKESNEFEVNLKDIETLGIKYDAVYTSYDPKIMEMAE 291
 233 YFAHEKVGTLIVRFDTPNPSKESNEFEVNLKDIETLGIKYDAVYTSYDPKIMEMAE 291
 233 YFAHEKVGTLIVRFDTPNPSKESNEFEVNLKDIETLGIKYDAVYTSYDPKIMEMAE 291
 292 SLILOGKAYIDTPPEOMKREKMDGIESCRNNTVEENLSLMKEVNTGEMGCCVRGK 351
 293 QIIXDGNAYADTEKEVNAEQPMNGKPSKRRDASVEENLAFEEKKGTPEGLRWCIRAK 352
 352 LDMODPNKSLRDPVYRCNTDPHHRVGSRYKVPYDACPFDVDALEGVTHALSSSEYHD 411
 353 MSVONPNKAMDPIVYRCNPAPHRHTGKMKIPIYDACPISVDSIEVTHALRTIERD 412
 412 RNAQYRILQDMGLRVEIYFSPRLNMYTLLSKRKLMPYQNKKEVDWTPRPTVQGI 471
 413 RNPQYQWFLDTLKLRLHVQIMDFARNFIRTLSSKRLTKLVQGVVGMGMDSPRFTIRGI 472
 472 VRBGLKVALIQLTIQOGASKNLNMEDKMTINKIIDIYVCARHTAVLKDQYIFTLT 531
 473 RRRGWTIPALREFFILKQSPSKNITNLDWTLWATNKXYIDPVAAPHRTAILKKDWKALVK 532
 532 NGPEEPFVRILPRHKKPFEGAGKATTPANRILMDVADAALINKGEVTLMDGNATVKEI 591
 533 GGPATPYIEEKPKKGNPAVGMKKVYFGNTVIFDQEDAKSKRQDEBITLMSGNATVKEI 592
 592 KVE--SGVITELVGLHLEGSVTKTKITWLA-DIEELVPLSLVEFDYLLSKKLEDE 648
 593 ETDPTSGIVKLELEHLEHEDGFKTEKKVYTWLSTEGQDLIIVELVDFYLLNKOTLQDD 652
 649 DFLNLNCTRREBIPALGDAMRNKIGETITQERKGYRCDAPITRSKSKVPLFAIPDG 708
 653 VLEVLNKNTEFRDADAVACVABLKEGDIIOFERKGYRVDRAVY-GKDAVALFINIPFG 711
 709 R 709
 712 K 712

RESULT 12
 AAU40163 standard; Protein; 581 AA.

AAU40163;
 13-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #1059.
 XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

XX 02-JUN-2000; 2000US-208841P.

XX 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 XX WPI; 2001-616774/71.

XX N-PSDB; AAS59510.

XX Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -

PS Example 1; SEQ ID No 1358; 1069pp; English.

CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 581 AA;

Query Match 18.7%; Score 697.5; DB 22; Length 581;
 Best Local Similarity 32.0%; Pred. No. 1.1e-52;
 Matches 186; Conservative 98; Mismatches 222; Indels 75; Gaps 17;

158 YKEALNE--VVAAPVGRKGIKSPAPSLKEKVDKSPAPAEVDLPGAKVGVCFAPAE 215
 4 YRDPPLDSGKVPAAVAPRG--NPAESSDPITHOVVADIQQ-DTYG--GRVGRPRPE 56
 216 PSGLYHIGHAKAALINKYFAERYQRLIVRFDTPNPSKESNEFEVNLKDIETLGIRYDA 275
 57 PNGYLHIGHAKAYITDGVADREPGTCNLRLDTPNPTETREYVESITLADIEMLGYSRAH 116
 276 VTYTSYDPKIMEMAESLILQKAYIDTPPEOMKREK---MDGISRCRNNTVEENLS 331
 117 VVHASDYFEQLYEWAQYLVGLAVYDDQSPETIREQGGYKGISSPYRNRPAAEESLN 176
 332 LKMEVNVGTERRGMCVRGLDMDQRPKSLRDPVYRCNTDPHHRVGSRYKVPYDACPISVDSIEVTHALRTIERD 391

```

Db 177 LIRMRAGEPFGDSRCLARIDMOMENWMLDPVWYRIRHOAHSTGTEKCIYPTDWAH 236
Qy 392 PFDVDALEGVTHALRSSEYHDNNAQYRILODMGLRVEI-----YEPSRLNMYTLLS 444
Db 237 GQSDAIEGVTHSLCSLEFNS-----HRLPYDMFSLHLPDGPAPKQREFALELTHITTS 291
Qy 445 KRKLMPQNKVQKVEDMTDPRPTVOGIVRGLKVEALIQFLQGCASKNLMEKDKMT 504
Db 292 KRLKSLVTNNIIVGMDPRMPTLGMRRRGYPAAIRAFQCAVGTTONSVKALIEPES 351
Qy 505 INKTIIDVCARHTAVLKDOQVI-----FTLTNGPEEPFRILPRHKPF 548
Db 352 FVRRELNATQRRNAVHLPLKVLVDGMPTDDGMPVWFQLVNPNP----- 400
Qy 549 EGAGKATTPANRIWLDYADA-----AINKEEYTLMDGNAIYK--IKVESGV 597
Db 401 -DDGTRRVPFTGELWEADDFREDPPRKFPILSPGHEVRLR--GAYLVATADVYKNPDTG 457
Qy 598 ITELVEGLHLEGSVYKT---KLKT--WLADIEELVLSLVEFDYILSKK-KLEBEDF 650
Db 458 IAEVHASVDPSQSGGTAPDGRKVSMTMHWVS-AGHAIVTANLYDRLSAIPSGQTGEA 516
Qy 651 LDNLNPCRREI-PALGDANMNIRKGEIILERKGYRCD 690
Db 517 LDDLNPHSRETLTEVMAEPALANVAPGSEVQPERLGYTPAD 557

```

RESULT 13

ABU06096 standard; Protein; 562 AA.

ID ABU06096;

AC ABU06096;

XX 22-JAN-2003 (first entry)

XX N. meningitidis vaccine antigen #72.

XX Vaccine; antigen; meningococcal disease; pathogenic bacteria;

XX meningitis.

XX Neisseria meningitidis group B.

XX WO200277648-A2.

XX 03-OCT-2002.

XX 22-MAR-2002; 2002MO-GB01399.

XX 22-MAR-2001; 2001GB-0007219.

XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.

XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

XX Robinson A, Goringe AR, Hudson MJ, Bracegirdle P, West DM;

XX Olliver KU, Kroll JS, Langford PR;

XX WPI; 2003-018958/01.

XX N-PSDB; ABX09930.

XX Identifying an antigen for manufacturing a vaccine against

XX meningococcal infection, comprises contacting antibodies with

XX polypeptides, detecting polypeptide-antibody complexes, and identifying

XX bound polypeptides as antigens -

XX Claim 37, Page 309-310; 310pp; English.

XX The invention relates to identifying an antigen comprising:

XX (a) obtaining antibodies against a commensal bacteria, or an extract

XX from a commensal bacteria; (b) contacting the antibodies with

XX polypeptides obtained from an expression library of either a commensal

XX or pathogenic bacteria; (c) determining whether the polypeptides bind to

XX antibodies; and (d) (where a polypeptide binds to an antibody)

```

CC identifying that polypeptide as an antigen. Also included are:
CC (1) a method of preparing a vaccine composition, comprising identifying
CC an antigen with the above method, and combining the antigen with a
CC carrier; (2) a vaccine composition obtained by the above methods;
CC (3) an antigen identified by the above methods; (4) a polypeptide
CC encoded by all or a part of a nucleic acid sequence comprising the
CC Neisseria lactamica DNA sequences detailed in the specification;
CC (5) a vector comprising the nucleic acid molecule; (6) a method of
CC preparing a composition for vaccination against infection by pathogenic
CC bacteria, comprising: (a) obtaining a first antigen from a commensal
CC Neisseria; (b) comparing the amino acid sequence of the first antigen
CC with the amino acid sequence of the second antigen from a pathogenic
CC bacteria, or comparing the sequence of a nucleic acid which codes for the
CC first antigen with the sequence of the nucleic acid that codes for the
CC second antigen; and if the first antigen is homologous to the second
CC antigen or if the nucleic acid sequence for the first antigen is
CC homologous to that of the second antigen, and (c) preparing a composition
CC for vaccination against bacterial infection comprising the first antigen;
CC and (7) an antibody that binds to the polypeptide antigen.
CC The method is useful in screening commensal and pathogenic
CC bacteria for previously unidentified vaccine antigens by identifying
CC polypeptide antigens that bind to sera raised against commensal
CC bacterial proteins. The polypeptide is useful as a vaccine antigen
CC which may be used in the manufacture of a medicament for vaccination
CC against meningococcal infection (e.g. meningitis). The present
CC sequence represents an antigenic protein from the pathogenic bacteria N.
CC meningitidis.
SQ Sequence 562 AA;

```

Query Match 17.8%; Score 666.5; DB 24; Length 562;

Best Local Similarity 34.7%; Pred. No. 66-50;

Matches 184; Conservative 84; Mismatches 217; Indels 45; Gaps 14;

Qy 197 EVDLPKAVGVKVCVRPAPEPSGVYLAIGHAKALINKFYAERYQRLIVRFDTPNPKESN 256

Db 18 BEDLESKHTAVQRRPPEPBGVLIHAKSICINPLGAVIYDGLCNLRFPDTPKEND 77

Qy 257 EIVENLKDIEITLGIKY-DAVYTSQYPPKIMENASLIKQKAYIDDTPEQNRKER-- 313

Db 78 EYVNAIKEDVWMLGFHNAAGEPRFASNYFDQLYDAVGLIKQKAYVDDLTPPEMRERYGT 137

Qy 314 --MDGISRCNNVNEENLSMKEMVNGTERGMCQCRGKLDMDQDPKSLADPYYRNT 371

Db 138 LTEAGKNSPYRDRVNEENLDFTRPKNGEPDGSKTRLKIDMAGSNIMMDPVIYRIR 197

Qy 372 DPHHRVGSKYVYPTYPFACPFVDALSGVTHALRSSEYHDNNAQYRILODM--GLRRV 428

Db 198 AHHNNTGDKCIYMYDTHTDISDAIGITHSLCTLFPEARPLVDCVLDNIPAPHATRP 257

Qy 429 EYEFSLNMYVYLLSKRKLMPYQNKVVEDMTDPRPTVOGIVRGLKVEALIQFLIQ 488

Db 258 RQYEFSLRELLYITTSRKLNQLVKEVSGMDPRMPTTSGMRRRGYTPGLRFAFRA 317

Qy 489 GASKNLNMEMDKMTINKKIIDVCAHRTAVLADQVIFLTN--GPEEPFRILPRH 545

Db 318 GISSENIIVDMSVLEGAIREELENSAPRLMVLNPLKV--TLTNFETGRTS--RRAAFH 373

Qy 546 KKEFGAGKATTPANRIWLDYADA-----AINKEEYTLMDGNAIYK--EIKVE 594

Db 374 PNHEMGREVPISQTIYIEADDFPAENPDKFKILRGEVRLRH--GYVAKCEVAVDE 431

Qy 595 SGVITTELVEGL-----HLEGSVYTKLKITWLADIEELVLSLVEFDYILSKKLE-- 645

Db 432 AGNVVELKCSIDHDTLGNPEG--RKVGVIHWVS--AEHAIEIKVRLDRLFTVERPDAV 488

Qy 646 --EBEDFL--DNANPCRREIPALGDANMNIRKGEIILERKGYRCD 690

Db 489 RGEQGEYLPFTDPLNPESVKEITAYAEPAADLPASRSKQPERIGYFVTD 538

RESULT 14

ABP77077

ID ABB77077 standard; Protein; 573 AA.
 XX ABB77077;
 AC
 XX
 DT 07-MAR-2003 (first entry)
 XX
 DE N. gonorrhoeae amino acid sequence SEQ ID 684.
 XX
 KM Antibacterial; infection; vaccine; gene therapy.
 XX
 OS Neisseria gonorrhoeae.
 XX
 PN W0200279243-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 12-FEB-2002; 2002WO-1B02069.
 XX
 PR 12-FEB-2001; 2001GB-0003424.
 XX
 PA (CHIR-) CHIRON SPA.
 XX
 PI Fontana MR, Pizza M, Maignani V, Monaci B;
 XX
 DR WPI; 2003-058415/05.
 XX
 DR N-PSDB; AB238047.
 XX
 PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 PT medicament for treating or preventing N. gonorrhoeae infection
 XX
 PS Disclosure; Page 239; 815pp; English.
 XX
 CC The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABB76736-ABP81046 represent nucleic acid
 CC molecules of the invention.
 CC
 SQ Sequence 573 AA;
 XX
 Query Match 17.7%; Score 661.5; DB 24; Length 573;
 Best Local Similarity 34.5%; Pred. No. 1.7e-49;
 Matches 183; Conservative 83; Mismatches 219; Indels 45; Gaps 14;
 QY 197 EVDLPKAVGVKVPAPAPSPGVLHIGAKAALINKYFAERYQRLIVFPDTPNPKESN 256
 DB 29 EDDLSKGHEAVQTRPFRPNQYLIHIGAKSICLNFGLAYIYDGLCNLRFDDTPNPKEND 88
 QY 257 EFVENLKDIEITLGIKY-DAVYTSDFPKLMEASLIIKQKAYIDDPKQNRKER- 313
 DB 89 EYVNAIKEDVEWMLGFHWAGBPRFASDYPDRLYDVAVGLIKQKAYVDLTPBEMREYRG 148
 QY 314 --MGISRCNNVTVEEVLSTMKEMVNTERGMOCCVARGKIDMODPKNSLADPPYIRGNT 371
 DB 149 LTERAGKSPYDRSIEENLDFTRMKNGSPDGSKTLMLKIDMAAGINMMDPVYIRLR 208
 QY 372 DPHRVGSKYKVVYTFYDFACFVDALLEGVTHALRSSEHDNNAQYRILQD--MGLRRV 428
 DB 209 AHHNNTGDKICITPMYDTHGICSDAIBGITHSLCTLEBARPLDYWDLIIIPALHATRP 268
 QY 429 EIVFSSRLNMYVYLLSKRKLMPVQNKVEDWTDPRPFTVOGIVARGLKEVALIQIPLIQ 488
 DB 269 RQYEFSSRLNMYVYLLSKRKLMPVQNKVEDWTDPRPFTVOGIVARGLKEVALIQIPL 328
 QY 489 GASKNLIMEMDKMTINKKIIDPVCARHTAVLQDQYIFLTLN---GPEEPFARILPRH 545
 DB 329 GISSEENLVDSVEGALRELENSAPRLMVLNPLKY-7LITNFOAKRTGS--RRAAFH 384
 QY 546 KKFPGAGKATTPANRIWLDYADAA-----AINKGEVTLMDMGNAIVK--EIKVE 594

DB 385 PNHEEMGDREVPVSQTIYIEADDFAEENPPKGFKRLIGCEVRLNH--GYVIKCEGVKDE 442
 QY 595 SGVITELVGEI-----HLEGSVKTKTKITWLDADIEELVPLSLVEPDYLSKCK---- 643
 DB 443 AGNVELKCSIDHDTLGKNPEG--RKVGVIHWVS-ABHAARIVRLYDLRFTYERGA 499
 QY 644 LEEDEDFL---DNLNPTRRRIEIPALGDANMKRNGEIIQLERKGYRCD 690
 DB 500 RGEDEGLPFTDFLNPESVKEITAVAPPAKADLPASRWQPERIGYFTVD 549
 RESULT 15
 AAB96155
 ID AAB96155 standard; Protein; 577 AA.
 XX
 AC AAB96155;
 XX
 DT 29-OCT-2001 (first entry)
 XX
 DE Putative P. abyssi glutamyl- and glutaminyl-tRNA synthetase.
 XX
 KM Hyperthermophilic archaeon; hyperthermophilic protein.
 XX
 OS Pyrococcus abyssi.
 XX
 PN FR2792651-A1.
 XX
 PD 27-OCT-2000.
 XX
 PF 21-APR-1999; 99FR-0005034.
 XX
 PR 21-APR-1999; 99FR-0005034.
 XX
 PA (CNRS) CNRS CENT NAT RECH SCI.
 PA (IFREMER) IFREMER INST FR RECH EXPL MER.
 PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
 PI Querellou J, Weissenbach J, Saurin W, Heilig R;
 DR WPI; 2001-126236/14.
 XX
 XX New nucleotide sequences isolated from Pyrococcus abyssi encode
 PT proteins useful in industry -
 XX
 PS Claim 7; Pages 793-794, 1657pp; French.
 XX
 CC The present invention relates to the genomic sequence of Pyrococcus
 CC abyssi (see AAF66431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
 CC a hyperthermophilic archaeon, which is isolated from deep-sea
 CC hydrothermal vents. The present sequence is one such P. abyssi protein.
 CC The proteins of the present invention have various potential industrial
 CC uses, since the proteins are stable at very high temperatures, some up to
 CC 110 degrees centigrade.
 CC Note: This patent is in the same patent family as W0200065062, which
 CC contains additional sequences as shown in AAB99132-AAB99143.
 CC
 AAH75903-AAH75920 and AAG66436.
 XX
 SQ Sequence 577 AA;
 XX
 Query Match 17.5%; Score 653; DB 22; Length 577;
 Best Local Similarity 30.8%; Pred. No. 9.9e-49;
 Matches 181; Conservative 92; Mismatches 252; Indels 62; Gaps 13;
 QY 131 IGQWESLRSKSK---YQNLVWFNSID-SEYKALNEVVAFAVGRGIGKSPAPSLKE 185
 DB 38 LGEHPELRSKAREIVPIINKVVEEVLSTLDEQAKLMEIY-----PEYFE 83
 QY 186 KVHDSKPSAPAEVDLPKAVGVKVPAPSPGVLHIGAKAALINKYFAERYQRLIVR 245
 DB 84 KKEEKKEKKGLPPLPAKGVKVTFRAPNPDDGFHGNARAAALISYVAKMGYKFTLR 143
 QY 246 FDDNPPKESNE--FVNLLKDIEITLGIKYDAVYTSDFPKLMEASLIIKQKAYIDD 303

Db 144 FDDTDPKVRPEPIFYEWIIEIDLMLGKPEIIVASDRLELYKAEELIKMGKAVVCT 203
QY 304 TPKEQMEKERMDGIESRCNNVTBENLSLMEKVNTERGMOCRCGLDMQDPKSLRD 363
Db 204 CKPEKFEHLKDXGIPCPHDEPVEQLEMRKMLNGEYKEGAVVRIKTDLNHPNPAVRD 263
QY 364 PYYRCNTDPHH-RVGSKYKVPYTFACPFVDALGVTALRSSEYHNRNAQYRIIJD 422
Db 264 WPAIRIVDNPNHPRAKNRYRWPLYNFASALDHGLVTHIFRQGEHLENETRQRIYEX 323
QY 423 MGLRVEIYEFSRLNMVYTLISKRLMFVNKKVEDTDPFPVTQGIVRGLKEALI 482
Db 324 FGWEYPTVHHGRLSIEGVILSKSTRKGIIEGKYLGMDDPRLGITIRALRRRGILPEAIK 383
QY 483 QFILOQASKRLNMEMDKLMTINKKIIDPVCARTAVLKDQVIFTLTNGPEEPFVRL 542
Db 384 ELIEVGLKSDATVSWDNLAIRKLVDPANRYFFVA--DPVMEVEGAPE--FIKI 439
QY 543 PRHKKFEGAGKKATTF--ANRIWLDYADAAINKGEEVTLMDWGNALVKEIKVESGVITE 600
Db 440 PLHPDHPERGTRELRFTPGKEIYVSKDDLKRGSPVRLKDLFNVIEIVEGK---IKA 496
QY 601 LVGELHLEGSVYKTKLKITWLAD--IBELVPLSLVEFDYLISSKKLEDEDDFLDNINPC 657
Db 497 KFSFEYEIARKNKMIMHWBEGRPCEVIIP---EGDELIVRKGLLEKD----- 543
QY 658 TRREIPALGDANMNIRKGEIIOLEKRYRCDAFIRSSKPVULFA 704
Db 544 -----ANVRAGEIIVQFERFGFVRIDK--IEGEKVVAIYA 575

Search completed: January 25, 2004, 17:09:46
Job time : 628 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2004, 16:58:20 ; Search time 71 Seconds
(without alignments)

426,088 Million cell updates/sec

Title: US-09-831-683B-10

Perfect score: 3734
Sequence: 1 MEALSFKSDSPISIICAA.....SSKRVLFAPIDGRQASLS 715

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCOTS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1506	40.3	1512	4	US-09-443-184-48
2	1437	38.5	1440	3	US-09-357-251-37
3	708.5	19.0	589	4	US-09-252-991A-30001
4	528	14.1	313	4	US-09-328-352-7010
5	212.5	5.7	516	4	US-09-198-452A-598
6	212	5.7	484	3	US-08-913-578-2
7	212	5.7	484	3	US-08-785-427-2
8	209	5.6	483	4	US-09-107-532A-5896
9	205.5	5.5	480	2	US-08-962-203-2
10	205.5	5.5	480	3	US-09-282-125A-2
11	205.5	5.5	480	4	US-09-273-142-2
12	203	5.4	501	4	US-09-134-001C-4115
13	199	5.3	348	2	US-08-844-153-2
14	193.5	5.2	287	4	US-09-328-352-7096
15	187	5.0	502	4	US-09-252-991A-26917
16	185.5	5.0	362	4	US-09-252-991A-26917
17	171	4.6	524	4	US-09-328-352-6089
18	133	3.6	1275	3	US-09-120-513-2
19	133	3.6	1275	3	US-09-450-105-2
20	132.5	3.5	313	4	US-09-328-352-5463
21	119.5	3.2	1627	4	US-09-328-352-6604
22	116	3.1	804	1	US-08-785-428-2
23	116	3.1	804	2	US-08-996-797-2
24	115.5	3.1	2662	4	US-09-595-684B-31
25	114	2.9	574	3	US-09-352-990-14
26	110	2.9	479	4	US-09-134-001C-4128
27	110	2.9	515	4	US-09-107-532A-5317

28	108.5	2.9	2172	1	US-08-611-107-31	Sequence 31, Appl
29	108	2.9	3959	2	US-08-970-269A-30	Sequence 30, Appl
30	108	2.9	3959	2	US-09-407-562-30	Sequence 30, Appl
31	106	2.8	466	1	US-08-785-066-2	Sequence 2, Appl
32	106	2.8	466	3	US-09-007-355-2	Sequence 2, Appl
33	106	2.8	466	3	US-08-913-489-2	Sequence 2, Appl
34	106	2.8	775	1	US-07-966-278-1	Sequence 1, Appl
35	106	2.8	775	1	US-08-424-921-1	Sequence 1, Appl
36	106	2.8	775	2	US-08-556-355A-1	Sequence 1, Appl
37	106	2.8	775	2	US-07-803-672A-1	Sequence 1, Appl
38	106	2.8	775	4	US-09-244-889A-1	Sequence 1, Appl
39	106	2.8	776	2	US-08-688-649-37	Sequence 37, Appl
40	106	2.8	2257	1	US-08-611-107-10	Sequence 10, Appl
41	106	2.8	2257	2	US-08-422-560A-10	Sequence 10, Appl
42	106	2.8	2257	3	US-08-468-793-10	Sequence 10, Appl
43	105.5	2.8	1683	3	US-08-755-587-183	Sequence 183, App
44	104	2.8	946	3	US-08-560-005-4	Sequence 4, Appl
45	104	2.8	946	3	US-09-418-540-4	Sequence 4, Appl

ALIGNMENTS

```
RESULT 1
US-09-443-184-48
; Sequence 48, Application US/09443184A
; Patent No. 6372431
; GENERAL INFORMATION:
; APPLICANT: Cunningham, Mary Jane
; APPLICANT: Zweigler, Gary
; APPLICANT: Kaser, Matthew R.
; APPLICANT: Panzer, Scott
; APPLICANT: Selhammer, Jeffrey J.
; APPLICANT: Yue, Henry
; APPLICANT: Baughn, Mariah
; APPLICANT: Azimzai, Yalda
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: MAMMALIAN TOXICOLOGICAL RESPONSE MARKERS
; FILE REFERENCE: PC-0007 US
; CURRENT FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PERL Program
; SEQ ID NO 48
; LENGTH: 1512
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6372431 2302721CD1
US-09-443-184-48
Query Match
Best Local Similarity 42.2%; Pred. No. 1.1e-143;
Matches 305; Conservative 146; Mismatches 232; Indels 40; Gaps 9;
4 ALSSKSDSPISIICAAQLVGLPTLINSGLAASAPTLQFASGSLGACVNPITLIYARCA 63
5 SLTNSGDPGPPGALLAAVHVVDVSI--SVBEGKENILHVSSENIPTDVNSIILYARVA 62
64 SIATSGKNDIEFGHVVEWLEAPTLISGSE--FENACLFVDCPLASRELVGHGLTIDI 122
63 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHNCJSLKTYLVGNSLIDL 122
123 AVSNLGIQGWESLRKSKKYONLVWFNSIDSEYKALNEVVAAPVGRGIGKSPAPS 182
123 CVMATLKGMAWQGLQKQKAPVHVKNMFGLEAM--QAFQSV-----G 164
183 LKERYHSDKPSAE-----VDLQKAGKGVCPAPPSGYLHGHAKALLNKYF 234
165 TKMNVSTTKAVAEKQDVGFPELPGAEWKGTVFPPRASGYLHGHAKALLNQHY 224
235 AERQGRHLIVAFDDTINSKESNEFEVNLKDIETLGIKYDAVTYTSDFPRLMEASLI 294
```

Dh	10	RVAPEKDDVGKVELPBGEMKGVTRPPEASGYLHGAKALLNQHVNFKGLIM	1.62
Qy	245	RFDDTTPRESNEPBNELNKDLETIGTKYDAVYTSDFPYFLYMASSLSLKOGAYVDDT	30.44
Dh	163	RFDDTTPREKEDDFEVLLEDAVMAHTRPDQFTYTSDFETIMYAKELIOEGAYVDDT	22.22
Qy	305	PKQMKERKEMDIESRCNNVTVEENLSLMKSMVNGTERGMOCYRGKLMODPNKSLRDP	36.44
Dh	223	PAQOMAEHQEIQIESHKNRPLEKXLOQWEEKKGSQGHSCSLRAKIDMSNNQCMRDP	28.22
Qy	365	VYRCHTDHHRVGSKKYKYPYDPACFPDNLBSGYTHALRSSYHNRNQAUYRILODMG	42.44
Dh	283	TLVRCKIOHPPTGKNYVPTYPDPACPIVDSIGVTHALRTEYHORDQFWIIIEALG	34.22
Qy	425	LRKEVEIESRNNVYTLTSKRLKLMFQONKKVEDWTDPRPPTQGIIVREGLKVEALIQF	48.44
Dh	343	IRKPEYIWEISRLNLTNNYLSKRKLWFLPNEGVLQGMDDPRPYTARGVLRRGMTVEGLKQF	40.22
Qy	485	ILOOGASKOLNIMENDKLTINKKIIDPYCARHVALKQORVITLLTNGEPEPRVILPR	54.44
Dh	403	IAAQGSSRSVVNMEDMKIWAFFKKVIDPVAARYVALAKKEVYI PVNVEAOE - MKEVAK	46.0
Qy	545	HKAPEGAGKATTPANRIMLDVADAAANKSEBVTLMDDMGALIKEI - KYESGYTELVG	60.30
Dh	461	HPKNEPEVGLKPYWYSKAPFIBGADLMTESBEBMTFLINMGMLNTKLNKADGKILSLDA	52.0
Qy	604	ELHLEGSVYTKLKITWLADIIEELVPLS - VEPDYLSKKKLEEDDFLNDLNPCTRR	66.1
Dh	521	KFLIEKDDKTKTKYTWLAETTHALPIRIVICVTEHLLTKFVLGDKEDDFKQVYKNKSHE	58.0
Qy	662	IPALGANNRRKRGIIOLERKGYVRD-----APRIRSKSVLYFAIPDG 708	
Dh	581	ELMLGPPCKLKDILKGDIIOLQRKGFIDQYIEPVSPIKSCAEPVULIYIPDG 633	

RESULT 3
US-09-252-991A-30001
; Sequence 30001, Application US/09252991A

```

: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196, 136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ. ID NOS: 33142
: SEQ ID NO 30001
: LENGTH: 589
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
: US-09-252-991A-30001

```

Query Match	15.0%;	Score 708.5;	DB 4;	Length 589;
Best Local Similarity	35.1%;	Pred. No. 4.8e-63;		
Matches 197;	Conservative 81;	Mismatches 217;	Indels 67;	Gaps 19
Qy	190	SKD-SAPE-----	YDLPGAKGKVCVRPAEDSGYLHGHAKAALLINKYFA	235
Db	30	TKDPMSPKPTAAANPLRQIVQADLDAGKAKAVITRFPPEBNGYLHGHAKSICLNFGIA		89
Qy	236	ERYGRLIVRPDDNPSKESNEFVENLIKDIETGIGKYDA-VYTTSDYFPFKLMEAEELI		294
Db	90	QEFAGDCHLRFDDNPAKEDDEYIDALEADIKMGFPQSGEVCYASNFDDLHAAWIELI		149
Qy	295	KQGAAYIDDPFKXEMRKER----MDGIESRCRNNTVEENSLTLMKEMVNGTERGMQCVRG		350
Db	150	KAGAAFAVCDGAGPEEMREYRGTLTTPGRNSPYRDSVSEENLDFAPMKAGPEPDGARSLRA		209

QY 351 KLDMDPNKSLRDPVYRCNTDPPHHRVSGKYKVFYTPYDPAFCFVADLGGVTHALSSSYH 410
DB 210 KIDMOSPMNNLRDPIYIRIRHAIHHQTGD KMCITYPSYDFTHQSDAIEGITHSICLTLEFE 269
QY 411 DRNMOYTHLDDMG-LRVEIYEFBSRLNMVYTLTKSLKLLMWONKRYEDWTDPFPYVQ 469
DB 270 DHRPLYLEFLANLPVPAQPROYEFERLNLNTVTSKRLKOLVDBGHSVGMDDPMSJLS 329
QY 470 GIVERGLKVEALIOFTLLOOGASKNLMLMEWDLMTINKKIIDPVCARHTAVLKQORVIFT 529
DB 330 GYRRRGYRPESIRNCEMIGVNRAGVYDIDGLESIDHDATAPRMCAUKPLKVV-- 387
QY 530 LTNGEPEEPVRI-LPRHKKFEGAGKATTPANRIMLYADAANA--KGEVYTL 580
DB 388 ITNYPEGVENLELPRHPR-EDMGVAVLPFGRELPIDAGDFEVPAGYKRLIPGGEVRL 446
QY 581 MDKGAIVK---EIVESGVITEL-----VGLHLEGSVYTKKLTWLADEELVP 629
DB 447 R-GSYVIRADEALDADGNIVELRCSYDPTLGK-NPEG--RKVKGVYHNV-PAEGSVE 500
QY 630 LSLVEFDYLI---SKKLEDEDFLDNLNP-----CTRREIPALGDAMNRNKRGEI 678
DB 501 CEVRLYDLRFSANPEKABEGGSFLDINADSLQVLAAG--RAEPLDQANPE-----DR 553
QY 679 TQERKGYRCDAPFIRSSKPV 700
DB 554 FOFEREGYFVADLKOSRPGKV 575

RESULT 4

US-09-328-352-7010
; Sequence 7010, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7010
; LENGTH: 313
; TYPE: PR1
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7010

Query Match 14.1%; Score 528; DB 4; Length 313;
Best Local Similarity 37.5%; Pred. No. 4,2e-45;
Matches 119; Conservative 55; Mismatches 113; Indels 30; Gaps 7;

QY 163 NEVVAAPVKGKIGKSPASLSKEKYHSDKSPAPV-----DLPGAIVKVC 209
DB 6 NDVVS-----NLPNRPPTPN-----THASVDSAQOQAGLDPVQRVITDDLAAGRAKQIV 55
QY 210 VRPAPEPGYLIHGAKKALLNKYFAERYOGRLIYRFPDTPNPKESNEFENLKDQIETL 269
DB 56 TRPPEPBGYLIHGIVKALICNPGVAEEFNGICNIRFPDTPDAEEQGYVDGIANVAKWL 115
QY 270 GIKYDA-VTYTSDYFPKLMEMASLTIKQAKAYIDTTPKORKEKEMDIE-----SRCRN 324
DB 116 GFSNNGEPRVAGYFDOLYAWAIOLEGGDAVYDQSPBELKLRGSEVVEGKNSPYDA 175
QY 325 TVEENLSMKEMVNGTERGMOCCVRGKLDMDPNKSLRDPVYRCNTDPPHHRVSGKYK 384
DB 176 SVEENLSMKEMVNGTERGMOCCVRGKLDMDPNKSLRDPVYRCNTDPPHHRVSGKYK 235
QY 385 PTYDFACFVVALGEGVTHALSSSEYHDNAQYRILDDMGARV-EIYEFBSRLNMVYTL 443
DB 236 PMYDAHPLSDAIEGITHSICLTLEFODHRPFYDWIVKYSKAVPHQYESRLNVDTYT 295
QY 444 SKRKLMLFVONKRYEDW 460
||||| : : : |

DB 296 SKRKLMLFVONKRYEDW-W 311

RESULT 5

US-09-198-452A-598
; Sequence 598, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffois, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 598
; LENGTH: 516
; TYPE: PR1
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-598

Query Match 5.7%; Score 212.5; DB 4; Length 516;
Best Local Similarity 25.6%; Pred. No. 1.6e-12;
Matches 97; Conservative 65; Mismatches 146; Indels 71; Gaps 16;

QY 208 VCVRPAPEPGYLIHGAKKALLNKYFAERYOGRLIYRFPDTPNPKESNEFENLKDQIETL 267
DB 17 VRVVAAPSPGDPVHGTAYMALFNEIFAKPKGKMLIRIEDTTRRQDYEENIFSLR 76
QY 268 TLGIKYDA-----VTYTSDFPKLME-MAESLTIKQAKAY-IDTTPKE--QMRKER-- 313
DB 77 WCGIOMBEGRVGVPGYRQOSERTKIYOGVETILTITDCAKGFATPDLAEBRAVA 136
QY 314 ---MDIESRCRNTVEENLSMKEMVNGTERGMOCCVRGKLDMDPNKSLRDPVYRCN 370
DB 137 LGYRGYDRRYRYSPE-----EVASREMAQGPYTR-----LKVPLSGECV 178
QY 371 -----TDPHHRVSGKYKVPYTPDFACRFPDALEGVTHALSSSYHNRNMOY 417
DB 179 FEDYSKRRVVPWADVDDQVLVSDGFPYHFAVVIDHLMGITHVLRGGEWLSSTPRL 238
QY 418 RILODMGLR-RVEIYEFBSRLNMVYTLTKSKR---LLMPVONKRYEDWTDPFPYVQI 472
DB 239 LLYAFGMEPVPFLMPLLNPDGTLKSKKNPSTIYYDSGYKE-AFVNPLTLMGYS 297
QY 473 RRG---LKYEALIOFTLLOOGASKNLMLMEWDLMTINKKIID---PVCARHTAVLKQ 524
DB 298 MEGDEEYVSLERITETENPRIGSGAVFDIQKLDMMNKYILNHEGSPCC---LLKEL 352
QY 525 RVITTLNGEPEEPVRI 543
DB 353 Q-GMLLN--DEFLKILP 367
||||| : : : |

RESULT 6

US-08-913-578-2
; Sequence 2, Application US/08913578
; Patent No. 6218159
; GENERAL INFORMATION:
; APPLICANT: Hodgson, John
; TITLE OF INVENTION: No. 6218159e1 tRNA synthetase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

```

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/913.578
  FILING DATE: 17-JAN-1997
  CLASSIFICATION: 435
  PRIORITY APPLICATION DATA:
    APPLICATION NUMBER: 9601069.9
    FILING DATE: 19-JAN-1996
    ATTORNEY/AGENT INFORMATION:
      NAME: Gimmil, Edward R.
      REGISTRATION NUMBER: 38,891
      REFERENCE/DOCKET NUMBER: P1352
      TELECOMMUNICATION INFORMATION:
        TELEPHONE: 610-270-4478
        TELEFAX: 610-270-5090
    TELETYPE:
      INFORMATION FOR SEQ ID NO: 2:
        SEQUENCE CHARACTERISTICS:
          LENGTH: 484 amino acids
          TYPE: amino acid
          STRANDEDNESS: single
          TOPOLOGY: linear
        MOLECULE TYPE: peptide
US-08-913-578-2

```

```

Query Match      5.7%; Score 212; DB 3; Length 484;
Best Local Similarity 20.8%; Pred. No. 1.6e-12;
Matches 106; Conservative 92; Mismatches 201; Indels 110; Gaps 18;

QY 207 KVCVPAPBEGSYHIGHAKAALLNKYAEYQGLIVRPDDTNPSSKESNEFVENLKDQI 266
DB 4 RIRVRYAPSPGYLHIGNARTALFNLYAKHYNGDFVIRIEDTKKNLEDEGETSQPDNL 63
QY 267 ETLGIXYD-AVTYSDYFPK-----LMEMASLHKGKAYIDDPKQMRKMDGI 317
DB 64 KWLGLDWDSEVDKONGPYRQSERQHIYQPLIDQLAEDKAYKCYWTEBELERERAOI 123
QY 318 ESRCRNNTVEENLSLWKEVNGTE-RGMQCCVRGKLDMDPNKSLRDPVYVRCNTDPHHR 376
DB 124 ARGEMPRYGGQHAHLTEBQROQFEABGRQPSIRFRVP-QNGTYSFDDWVKNISFD-SNG 181
QY 377 VGS---KYKVPYDPACPFVDALLEGVTHALRSSEYHNRNAQYRILQDMGLRRVEIYE 432
DB 182 IGDWVIYVKDGIPTYNFAVALIDHYMQISDVRIGDDHISNTPKQIMYEARGW--EPPR 238
QY 433 FSRLLMYVT---LISKR--KLWLVQNKVYEDTDRPFRPVQGIIVRGKLVKVEALIQFIL 486
DB 239 FGHMSLVNERKKLSKRDGQILOP-----IEQYRD-----LGYLPEALFNPIA 282
QY 487 QQGASKNLNLMEMWDLWTINKKIIDPVCARHTAVLKDQRVIFTLTNGPEEPFVRILPRHK 546
DB 283 -----LIGWSP-----EGEBEIFS-----KEEFIKIFDEKR 308
QY 547 KPEGAKKATTPANRIMLDVADAAAIKNGEEVTLMDGNALVKEIKVESGYITELVGLH 606
DB 309 L-----SKSPAFPDQKQLAWNNQYMKQDTEYFQ---LALPHLIKANLIPVPSBED 359
QY 607 LEGSVKTTKLTITWLADIIEVLVPLSVFVDYLSKKLEBDEDFLDNINPCTREIRIPLG 666
DB 360 LSWGRKLIALYOKENSYAGEIVPLSEMPF-----KEMPALG 395
QY 667 DANMNRINRGELIQERKGYRCA--PF 693
DB 396 EBEQVINGEQVPELMTLHFLSKLALBEP 424

```

RESULT 7
 US-08-785-427-2
 Sequence 2, Application US/08785427
 Patent No. 6238900
 GENERAL INFORMATION:

```

APPLICANT: Hodgson, John
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 6238900e1 crna synthetase
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
  ADDRESSEE: SmithKline Beecham Corporation
  STREET: 709 Swedeland Road
  CITY: King of Prussia
  STATE: PA
  COUNTRY: USA
  ZIP: 19406-0939
COMPUTER READABLE FORM:
  MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/785.427
  FILING DATE: 17-JAN-1997
  CLASSIFICATION: 435
  PRIORITY APPLICATION DATA:
    APPLICATION NUMBER: 9601069.9
    FILING DATE: 19-JAN-1996
    ATTORNEY/AGENT INFORMATION:
      NAME: Gimmil, Edward R.
      REGISTRATION NUMBER: 38,891
      REFERENCE/DOCKET NUMBER: P1352
      TELECOMMUNICATION INFORMATION:
        TELEPHONE: 610-270-4478
        TELEFAX: 610-270-5090
    TELETYPE:
      INFORMATION FOR SEQ ID NO: 2:
        SEQUENCE CHARACTERISTICS:
          LENGTH: 484 amino acids
          TYPE: amino acid
          STRANDEDNESS: single
          TOPOLOGY: linear
        MOLECULE TYPE: peptide
US-08-785-427-2

```

```

Query Match      5.7%; Score 212; DB 3; Length 484;
Best Local Similarity 20.8%; Pred. No. 1.6e-12;
Matches 106; Conservative 92; Mismatches 201; Indels 110; Gaps 18;

QY 207 KVCVPAPBEGSYHIGHAKAALLNKYAEYQGLIVRPDDTNPSSKESNEFVENLKDQI 266
DB 4 RIRVRYAPSPGYLHIGNARTALFNLYAKHYNGDFVIRIEDTKKNLEDEGETSQPDNL 63
QY 267 ETLGIXYD-AVTYSDYFPK-----LMEMASLHKGKAYIDDPKQMRKMDGI 317
DB 64 KWLGLDWDSEVDKONGPYRQSERQHIYQPLIDQLAEDKAYKCYWTEBELERERAOI 123
QY 318 ESRCRNNTVEENLSLWKEVNGTE-RGMQCCVRGKLDMDPNKSLRDPVYVRCNTDPHHR 376
DB 124 ARGEMPRYGGQHAHLTEBQROQFEABGRQPSIRFRVP-QNGTYSFDDWVKNISFD-SNG 181
QY 377 VGS---KYKVPYDPACPFVDALLEGVTHALRSSEYHNRNAQYRILQDMGLRRVEIYE 432
DB 182 IGDWVIYVKDGIPTYNFAVALIDHYMQISDVRIGDDHISNTPKQIMYEARGW--EPPR 238
QY 433 FSRLLMYVT---LISKR--KLWLVQNKVYEDTDRPFRPVQGIIVRGKLVKVEALIQFIL 486
DB 239 FGHMSLVNERKKLSKRDGQILOP-----IEQYRD-----LGYLPEALFNPIA 282
QY 487 QQGASKNLNLMEMWDLWTINKKIIDPVCARHTAVLKDQRVIFTLTNGPEEPFVRILPRHK 546
DB 283 -----LIGWSP-----EGEBEIFS-----KEEFIKIFDEKR 308
QY 547 KPEGAKKATTPANRIMLDVADAAAIKNGEEVTLMDGNALVKEIKVESGYITELVGLH 606
DB 309 L-----SKSPAFPDQKQLAWNNQYMKQDTEYFQ---LALPHLIKANLIPVPSBED 359
QY 607 LEGSVKTTKLTITWLADIIEVLVPLSVFVDYLSKKLEBDEDFLDNINPCTREIRIPLG 666

```

Db 360 LSWGRKLTALYQKMSYAGEIIVPLSEMF-----KEMPALG 395
 Qy 667 DANNENIRGEIIQERKGYRCDA--PF 693
 Db 396 EEOOVINGEQVPELMTHLFSLKEALBP 424

RESULT 8

US-09-107-532A-5896

/ Sequence 5896, Application US/09107532A
 / Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: LYNN A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESS: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneka

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781) 893-5007

TELEFAX: (781) 893-8277

INFORMATION FOR SEQ ID NO: 5896:

SEQUENCE CHARACTERISTICS:

LENGTH: 483 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...483

SEQUENCE DESCRIPTION: SEQ ID NO: 5896:

US-09-107-532A-5896

Query Match 5.6%; Score 209; DB 4; Length 483;
 Best Local Similarity 24.6%; Pred. No. 3.2e-12;
 Matches 78; Conservative 45; Mismatches 100; Indels 94; Gaps 13;

Qy 210 VRAPSPGVLHGHAKAALLINKYFAERYQRLIVRFDTPNPSKESNEFVENLKDIETL 269
 Db 3 VRAVSPGVLHGHAKAALLINKYFAERYQRLIVRFDTPNPSKESNEFVENLKDIETL 62

Qy 270 GIKXD-ATYTSVDP-----KLEMAESLIQ-----GKAYIDDPKQEMKER----- 313

Db 63 GMMNDESENPGEYGYRQSRKEIYQPLIDOLASNAKYCYCTEELERERQARG 122

Qy 314 -MDGIESCRNNVTYENISLWKE-----MVNGTERGMQCCVAGKLDMDPNKSL 361

Db 123 EMPHYAGTCA/LTPEEQAE--KEAQGLSESVFRFPNRTETAFTDMVKALISPSDN--- 177
 Qy 362 RDPVYACNTDPPHRRVSGKYK-----YPTDFACPFVDLLEGVTHALRSSEYDRNAQY 416
 Db 178 -----IGDFVIQKRDGMPTNFVAVDDHLMKTIHVLRGDD-HIANTP- 220
 Qy 417 YRIQDMGLRVEIYEFSSRLNMVYTLSSKRLMFWONKKEVDWTDPRFPVQGIY--RR 474
 Db 221 -----KQLMYE-----AFGWTPTPTFGMTLINSIT 247
 Qy 475 GLKV---EALIQITLQ 487
 Db 248 GKLSKRDESILQFIEQ 264

RESULT 9

US-08-962-203-2

/ Sequence 2, Application US/08962203

/ Patent No. 5976840

GENERAL INFORMATION:

APPLICANT: Jaworski, Deborah

APPLICANT: Lawlor, Elizabeth

APPLICANT: Wang, Min

TITLE OF INVENTION: NOVEL STREPTOCOCCAL BRS

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESS: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: IBM Compatible

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/962,203

FILING DATE: 31-OCT-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/844,153

FILING DATE: 18-APR-1997

APPLICATION NUMBER: 9607992.6

FILING DATE: 18-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Gimmi, Edward R

REGISTRATION NUMBER: 38,891

REFERENCE/DOCKET NUMBER: P31459-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-4478

TELEFAX: 610-270-5090

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 480 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-962-203-2

Query Match 5.5%; Score 205.5; DB 2; Length 480;
 Best Local Similarity 19.1%; Pred. No. 7.1e-12;
 Matches 99; Conservative 82; Mismatches 170; Indels 167; Gaps 19;

Qy 210 VRAPSPGVLHGHAKAALLINKYFAERYQRLIVRFDTPNPSKESNEFVENLKDIETL 269
 Db 1 VRAVSPGVLHGHAKAALLINKYFAERYQRLIVRFDTPNPSKESNEFVENLKDIETL 60

Qy 270 GIKYDAVTYT-----SDYPKLMEASLIQKQKAYIDDPKQEMKER----- 313

```

Db      61 GMDWDESPESHENYRQSERLDLYQKTIQDLAEGKAYKSYTTEBELAERERQEAETP 120
Qy      314 -----MDGIESRCNNNTVEENLSMKEMVNGTERGMOCVGRK 351
Db      121 RYNEYLGMSEBEKAAYIAEREAAGIIPVRLAVNESGIYKWHDM-----VKGD 169
Qy      352 LDMQDPNKSLLDPVYRCNTDPHHRVGSKTV-----YPTDFACFPVDALGVTALRS 406
Db      170 IEFGGN-----IGGDWVIQKKDGYPTNFVAVIDHDMQISHVIRG 211
Qy      407 SEYHDRAQYRILQDMGLRRVEIYEFSLNMVYT-----LLSKR--KLWFOVKKVED 459
Db      212 DDHIANTPKQMLMVEALGM---EAPFEGHMTLIINSETGKLSKRDNTLTQFIEDYR--- 265
Qy      460 WTDPRFPVQGIVRGKVEALLQFILOOGASKNLNMEMDKLMTINKKIIDPVCARHTA 519
Db      266 -----KKGVLPEAVFNFI-----LLGM-----NFGGDEIFSR--- 294
Qy      520 VLKQORVIFLTNGPEEPFVRILPRHKKFEGAGKATTFANRIMWDYADAAAIKGEVY 579
Db      295 -----BELIKLPDENR-----LSKSPAAFQDK-KLDMWSNDYIKNADLET 333
Qy      580 LMDGNAIVKEIKVESGITELVGEI-----HLEGVKTTLKITWLADEEL----- 627
Db      334 IFE---MAKPFLEAGRLTDKAELFDLYPKQMSVDEIIPLTDLFFSDPELLEARE 389
Qy      628 -----VPLSLVEFDYLISSKKLE---EDEDFLDNLP 656
Db      390 VMTGETVPTVLEAF-----KAKLEAMTDEFTENIFF 422

```

RESULT 10

```

US-09-282-125A-2
; Sequence 2, Application US/09282125A
; Patent No. 6165760
GENERAL INFORMATION:
APPLICANT: Jaworski, Deborah
Lawlor, Elizabeth
Mang, Min
TITLE OF INVENTION: NOVEL STREPTOCOCCAL ERS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESSES:
ADDRESSER: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/282,125A
FILING DATE: 07-Aug-2000
CLASSIFICATION: <Unknown>
18-APR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/962,203
FILING DATE: <Unknown>
APPLICATION NUMBER: 9607992.6
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimm, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31459-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX: <Unknown>
18-APR-1996
INFORMATION FOR SEQ ID NO: 2:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-282-125A-2

```

```

Query Match 5.5%; Score 205.5; DB 3; Length 480;
Best Local Similarity 19.1%; Pred. No. 7.1e-12;
Matches 99; Conservative 82; Mismatches 170; Indels 167; Gaps 19;

```

```

Qy      210 VRAPSPGVIHTHGAALINKYFAERYQRLIVREDTPNPSKESNEPVENLLKDIETL 269
Db      1 VRAPSPGLHIGNARTALFNLYAHNGGTFLIRIEDTRKHVDEGERSQLENLRL 60
Qy      270 GIKYDAVYTT-----SDYFPKLMEMASLIKQKAYIDTPKPKRMER----- 313
Db      61 GMDWDESPESHENYRQSERLDLYQKTIQDLAEGKAYKSYTTEBELAERERQEAETP 120
Qy      314 -----MDGIESRCNNNTVEENLSMKEMVNGTERGMOCVGRK 351
Db      121 RYNEYLGMSEBEKAAYIAEREAAGIIPVRLAVNESGIYKWHDM-----VKGD 169
Qy      352 LDMQDPNKSLLDPVYRCNTDPHHRVGSKTV-----YPTDFACFPVDALGVTALRS 406
Db      170 IEFGGN-----IGGDWVIQKKDGYPTNFVAVIDHDMQISHVIRG 211
Qy      407 SEYHDRAQYRILQDMGLRRVEIYEFSLNMVYT-----LLSKR--KLWFOVKKVED 459
Db      212 DDHIANTPKQMLMVEALGM---EAPFEGHMTLIINSETGKLSKRDNTLTQFIEDYR--- 265
Qy      460 WTDPRFPVQGIVRGKVEALLQFILOOGASKNLNMEMDKLMTINKKIIDPVCARHTA 519
Db      266 -----KKGVLPEAVFNFI-----LLGM-----NFGGDEIFSR--- 294
Qy      520 VLKQORVIFLTNGPEEPFVRILPRHKKFEGAGKATTFANRIMWDYADAAAIKGEVY 579
Db      295 -----BELIKLPDENR-----LSKSPAAFQDK-KLDMWSNDYIKNADLET 333
Qy      580 LMDGNAIVKEIKVESGITELVGEI-----HLEGVKTTLKITWLADEEL----- 627
Db      334 IFE---MAKPFLEAGRLTDKAELFDLYPKQMSVDEIIPLTDLFFSDPELLEARE 389
Qy      628 -----VPLSLVEFDYLISSKKLE---EDEDFLDNLP 656
Db      390 VMTGETVPTVLEAF-----KAKLEAMTDEFTENIFF 422

```

RESULT 11

```

US-09-273-142-2
; Sequence 2, Application US/09273142
; Patent No. 6300119
GENERAL INFORMATION:
APPLICANT: Jaworski, Deborah
Lawlor, Elizabeth
Mang, Min
TITLE OF INVENTION: NOVEL STREPTOCOCCAL ERS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESSES:
ADDRESSER: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/09/273.142
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/962.203
FILING DATE:
APPLICATION NUMBER: 9607992.6
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimm, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31459-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-273-142-2

Query Match 5.5%; Score 205.5; DB 4; Length 480;
Best Local Similarity 19.1%; Pred. No. 7,1e-12;
Matches 99; Conservative 82; Mismatches 170; Indels 167; Gaps 19;

QY 210 VRAPESGYLHIGAKAALLNKYFAERYGRLIVRPDTPNSKESNFVNLKDIETL 269
DB 1 VRAPESGYLHIGAKAALLNKYFAERYGRLIVRPDTPNSKESNFVNLKDIETL 60
QY 270 GIKDAVYTT-----SYPRKLMEMASLIKQKAYIDTPKQMKR----- 313
DB 61 GMDWSESHENYRQSERLDLYOKYIDQLLAEGKAYSYTEBELAERROEAAGETP 120
QY 314 -----MDGIESRCRNNTVEENLSLKEVNVGTERGMOCCVAGK 351
DB 121 RYINELGMSSEBEKAYIAERAGLIPTVRLAVNESIYKMDV-----VKGD 169
QY 352 LDMODPNKSLDPVYYRCNTDPHHRVSGKYV-----YPTVDFACPFVDALEGVTHALRS 406
DB 170 IEFEGN-----IGGDWVIQKDGQPTYNFAVIDDHMQSHVIRG 211
QY 407 SEYDRNAQYRIIDMGRLRVEIYESRLNMVTT-----LISKR--KLIMFVQKKVED 459
DB 212 DDHIANTPKQMLVVEALGM---EAPFEGHMTLIINSETKKLSKRDNTLTQFIEDYR--- 265
QY 460 WTDPRFTVQGIIVRGKLVKVEALQIFILQOGASKNLIMBWDKMTINKKIIDPVCAHRTA 519
DB 266 -----KKGYPBEAVFNFTA-----LIGM-----NFGGEDELFSR--- 294
QY 520 VLKQDQVIFTLTNGPEEPFVRILPRHKKPEGAKKATTFANRIMLDVADAANKGEEVT 579
DB 295 -----BELIKLPDENR-----LSKSPAFDQK-KIDMMSNDYIKKADLET 333
QY 580 LMDGNALVKEIKVESGVITELVGL-----HLEGVKTTLKITWLADIEEL----- 627
DB 334 IFE---MAKPFLEAGRLTDKAEKLFPLVYRQMKSVDEIIPLDLFFSDPELLEARE 389
QY 628 -----VPLSLVEPDYLSKKKLE---EDDFLNDLNP 656
DB 390 VMGGEVFTVLEAF-----KAKLEAMTDEFTVENIFP 422

RESULT 12
US-09-134-001C-4115
Sequence 4115, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134.001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4115
LENGTH: 501
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4115

Query Match 5.4%; Score 203; DB 4; Length 501;
Best Local Similarity 20.0%; Pred. No. 1.4e-11;
Matches 97; Conservative 86; Mismatches 183; Indels 118; Gaps 18;

QY 202 GAKV--GKVCVRFAPESGYLHIGAKAALLNKYFAERYGRLIVRPDTPNSKESNFV 259
DB 14 GVKMSRIRIVRYAPSTGYLHIGAKAALLNKYFAERYGRLIVRPDTPNSKESNFV 73
QY 260 ENLLKDIETIGIKYD-AVYTSYFP-KLMEMAE-----SLIKQKAYIDTPKQMR 310
DB 74 SSQFDNLKMLGLDWDSDVDKQFGFQSERAEITNPLIQLLBEDKAYCWTBEELE 133
QY 311 KER-----MGIESRCRNNTVEENLSLKEVNVGTERGMOCCVAGK 342
DB 134 AERBAQIARGMPRYGGQAHLTBEORQYEAERKPSIRFVXPQDQTYFENDM----- 187
QY 343 GMOCCVAGKLDMDPNKSLDPVYYRCNTDPHHRVGS---KYKVYPTVDFACPFVDALE 398
DB 188 -----VKGESIFSEDN-----IGMWIVKDGQPTYNFAVIDDHMT 224
QY 399 GVTHALRSSEYDRNAQYRIIDMGRLRVEIYESRLNMVTT-----LISKR--KLIMFV 452
DB 225 QISVINGDDHVSNTPRQMLIYEAFGW---EAPFEGHMTLIINSETKKLSKRDQIIQFI 281
QY 453 QNKVEVWTD--RFTVQGIYVRG---LKVEALQIFILQOGASKNLIMBWDKMTI 505
DB 282 EOVYDLGLPBEALFNFTLLGWSPEGEIEFSKEEPIKIDPEKLSKSPMFDROKLAW 341
QY 506 NKKIIDPVCAHRTAVLKDQVIFTLTNG-----PEEPFVRILPRHKKPEGAKKAT 557
DB 342 NNQYMK-----TKDTETVELALPHILKANLIPENPSEKOREMGRKLIAYOKEMS 392
QY 558 FANRI-----WLDYADAANKGEEVTIMDGNALVKEIKVESGVITELVGLHLEGSVKT 613
DB 393 YAGSIVLSEMFHEHMELEKDBQEVQ--GEQVPE-----LMMHLVKGKLSLSBEFA 443
QY 614 TKLK 617
DB 444 TEIK 447

RESULT 13
US-08-844-153-2
Sequence 2, Application US/08844153
Patent No. 5958734
GENERAL INFORMATION:
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 5958734el Compounds
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19046
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette

Query Match	5.3%	Score 199	DB 2	Length 348
Best Local Similarity	20.3%	Pred. NO. 1.9e-11		
Matches 67; Conservative	56;	Mismatches 109;	Indels 98;	Gaps 9;

RESULT 14

US-09-328-352-7096

RESULT 15
US-09-252-991A-28736

US-09-252-991A-28736

QY 425 --LRRV

Db 301 TLAEHIEHFDLSRVSLGGPIFDLEKLSWLNQGMIREQSVBEPAREVOKMALNPEYLMKIA 360
Qy 466 PTVOGIVRRGLKVEALIOFILOQASKNLNMENDKLMWTINKKIDPVCARHTAVL 521
Db 361 PHVQGRVENFSGIAPLAGFFPSGSGVPLDASLFEHKKL-----DPTQVROVLOL 408

Search completed: January 25, 2004, 17:15:58
Job time : 72 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2004, 17:09:55 ; Search time 247 Seconds

(without alignments)
598.448 Million cell updates/sec

Title: US-09-831-683B-10

Perfect score: 3734

Sequence: 1 MEALSRKSPISIIICAA.....SSKPVLPALPDGRQASLS 715

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 777136 seqs, 206736638 residues

Total number of hits satisfying chosen parameters: 777136

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1574.5	42.2	716	US-10-369-493-2175	Sequence 2175, Ap
2	1562	41.8	724	US-10-369-493-21897	Sequence 21897, A
3	1537	41.2	725	US-10-032-585-7526	Sequence 7526, Ap
4	1471.5	39.4	715	US-10-128-714-3375	Sequence 3375, Ap
5	1465.5	39.2	715	US-10-128-714-8375	Sequence 8375, Ap
6	1459.5	39.1	606	US-10-369-493-42310	Sequence 4230, Ap
7	1349	36.1	618	US-10-369-493-52210	Sequence 5221, Ap
8	1339	35.9	1149	US-10-369-493-3100	Sequence 3100, Ap
9	1287.5	34.5	717	US-10-104-047-3100	Sequence 3100, Ap
10	753	20.2	554	US-10-369-493-751	Sequence 751, App
11	710.5	19.0	537	US-10-369-493-9677	Sequence 9677, App
12	707	18.9	554	US-10-369-493-21266	Sequence 21266, A
13	704.5	18.9	554	US-10-369-493-13759	Sequence 13759, A
14	703	18.8	554	US-10-369-493-412	Sequence 412, App
15	684.5	18.3	573	US-10-369-493-9393	Sequence 9393, Ap

16	683.5	18.3	580	US-10-369-493-17582	Sequence 17582, A
17	674.5	18.1	573	US-10-369-493-9215	Sequence 9215, Ap
18	672.5	18.0	551	US-10-369-493-19380	Sequence 19380, A
19	672.5	18.0	569	US-10-369-493-8623	Sequence 8623, Ap
20	662.5	17.7	574	US-10-078-770-166	Sequence 166, App
21	658	17.6	556	US-10-369-493-7538	Sequence 7538, App
22	658	17.6	573	US-10-369-493-4778	Sequence 4778, Ap
23	653	17.5	571	US-10-369-493-21578	Sequence 21578, A
24	649	17.4	542	US-10-369-493-20659	Sequence 20659, A
25	624	16.7	570	US-10-369-493-1319	Sequence 1319, Ap
26	624	16.7	570	US-10-369-493-20384	Sequence 20384, A
27	589.5	15.8	553	US-10-369-493-1067	Sequence 1067, Ap
28	588	15.7	799	US-10-032-585-7457	Sequence 7457, Ap
29	573	15.3	786	US-10-369-493-6228	Sequence 6228, Ap
30	572	15.3	811	US-10-369-493-32597	Sequence 32597, A
31	561.5	15.0	775	US-10-153-668-300	Sequence 300, App
32	559	15.0	551	US-10-369-493-930	Sequence 930, App
33	537.5	14.4	553	US-10-369-493-1102	Sequence 1102, App
34	518.5	13.9	562	US-10-369-493-11272	Sequence 11272, A
35	480.5	12.9	550	US-10-369-493-18217	Sequence 18217, A
36	477.5	12.8	544	US-10-369-493-22958	Sequence 22958, A
37	464	12.4	540	US-10-369-493-13582	Sequence 13582, A
38	437	11.7	541	US-10-369-493-10980	Sequence 10980, A
39	424	11.4	433	US-10-078-770-164	Sequence 164, App
40	381.5	10.2	586	US-10-369-493-18607	Sequence 18607, A
41	364	9.7	383	US-10-078-770-170	Sequence 170, App
42	271	7.3	473	US-10-369-493-92	Sequence 92, App
43	255	6.8	469	US-10-369-493-3022	Sequence 3022, App
44	247	6.6	487	US-10-369-493-3071	Sequence 3071, Ap
45	243	6.5	463	US-09-815-242-11325	Sequence 11325, A

ALIGNMENTS

RESULT 1
US-10-369-493-2175
Sequence 2175, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 2175
LENGTH: 716
TYPE: PRT
ORGANISM: Schizosaccharomyces pombe
US-10-369-493-2175

Query Match 42.2%; Score 1574.5; DB 12; Length 716;
Best Local Similarity 49.5%; Pred. No. 3.8e-139;
Matches 319; Conservative 102; Mismatches 201; Indels 23; Gaps 10;
QY 70 GKNDIEFGHVEWLEVAFTPLSGS---EFENACLFDVDFGLASTFLVHGTLTADIAVW 125
DB 81 GENDRSL--VESWETA-SALAGHNFLELSLAQDDHLIMRSLFVGYSLTADPSIWM 137
QY 126 SNLAGIQREBSLRKSKYQYVRFWFSIDSEYKALNEVVAAPVGRGICKSPAPLKE 185
DB 138 GALKSNMAAGAVR-TQYFNLAWYKFDQS--NAVSVMEEFTKAVNISK-----KQ 188
QY 186 KVDHSKPSAPEVDLPQAKYGVKVCVRPAPPSGLTHGHAKAALINKYFAERYGRLIVR 245

Db 189 K---SSGPNY-EIGLPDAIDGKVTRPPEPSGYLIGHAKAALLNQYFANKYHGXILVR 244

Qy 246 FDDTNSKESNEFEVNLKDIETLGIKYDAVYTSDFPKLMEASLIGKQKAYIDTP 305

Db 245 FDDTNSKESNEFEVNLKDIETLGIKYDAVYTSDFPKLMEASLIGKQKAYIDTP 304

Qy 306 KEQKRRKRMQGISRCNNVTENLSLMKENVNGTERGMOCCVGLKMDOPNKSILRDPV 365

Db 305 VETRRHRTBGIPSKHRDRIEESLEIISMDKSDVGLKNCIRAKTSYENPNAMRPV 364

Qy 366 YRRCNTDPHHRVGSKYVPTYPDFACPFVDALLEGYTHALRSSEYHNRNAYRLLIDMGL 425

Db 365 IYRNLPHRHTGKRYAPFYDFACPIYDLSLEGYTHALRTEYRDRNPLQOMIKANL 424

Qy 426 RREIYEPSRLNMYTLLSKRKLIMFVONKVEWTDPRPTVOGIVRGLKVEALLQFI 485

Db 425 RKIHWFEFSRNPFRLLTSKRKLTEIVDHGLVWGMDDRPFTVGRVRRGKMTIEALQYI 484

Qy 486 LQOGASKNLNMEMDKLMTINKKIIDPYCARTAVLKNQRYIFTLTNGPEEPFVRLPRH 545

Db 485 VSQSPSKNILLDMTSEFATNKKIIDPVAPRHVAESGDVVKATIVNGPAAPYADRRPR 544

Qy 546 KKEFGAGKATTFANRIWLDYADAANKGSEVTLMDGNALVKEIKVE-SGVITELVGE 604

Db 545 KKNPELGKKSIFANEILIEGADAGSPKQDEBVTLMGKNAYREINRDSAGKYSKLJE 604

Qy 605 LHLEGSVTKTKLTWLDADIELVPLSIVEFDYLSKKKLEDEDFLDINPCTRRREIPA 664

Db 605 LHLGDPEKTEKTKYTLADTEDKTPVDLVDYLLITKOKLEGEYKDFLPPQTEFHSFV 664

Qy 665 LCGDANMNIRKGEIIOLEKGYRCDAPFIRSSKPVULFAIPDR 709

Db 665 PADGINKLKKGDIIQVERKGYIVDVPF-DGTQAVLENIPDGK 707

RESULT 2

US-10-369-493-21897

Sequence 21897, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT FILING DATE: 2003-02-28

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 21897

LENGTH: 724

TYPE: PRT

ORGANISM: Saccharomyces cerevisiae

US-10-369-493-21897

Query Match 41.8% Score 1562; DB 12; Length 724;

Best Local Similarity 45.7% Pred. No. 5.9e-138; Indels 56; Gaps 14;

Matches 328; Conservative 121; Mismatches 212;

Qy 27 LTINHSLAGSAPTLQFASGESLHGVN-----PIIYIRGASIASLGSKNIDIEFG-- 77

Db 21 LTIN-----GKAPVAVAEILAAKRYVNLAPNSIAIKLVDDKAPAAKLDATDETVFKI 75

Qy 78 -----HYEWLEAYAPTL--SGSEFENACLFPVDFGLASRTFLVGHGL--T 118

Db 76 TSKPAALFDNDGKQVAKVMVLAQKELVTKNPAKLSQSLFTLDSQNLNRTFLIG-GLKYS 134

Qy 119 IADIAVSNLAGIQGRWESLAKSKKYONLVVWENSID-----SEKELNEVVAAFVGR 173

Db 135 AADVACMGALRSNGM-CGSIHKONKVDVNVSRWYTLLEMDPIFGEAHDFLSKLELKSA 193

Qy 174 GIGSPAPSLKEKHYHDSKDSAPVVDLPGAKVGVYCRFAPBEGGYLHIGAKAALLNKY 233

Db 194 NVGR-----KKEHKNANF-----EIDLPAKMGVWVRFPPEPSGYLIGHAKAALLNQY 243

Qy 234 FAERYOGRILYRFPDTPNSKESNEFEVNLKDIETLGIKYDAVYTSDFPKLMEASL 293

Db 244 PAQVYKGLILIRFPDTPNSKEKEFEQSILEDLGLIGKDRITYSSDYFQEMWDYCYQM 303

Qy 294 IKQKAYIDTPPKQMKERMDGIESRCNNVTENLSLM-KENVNGTERGMOCCVGRKL 352

Db 304 IKQKAYCDDTPPKQMBERRMDGVAASARDRSVEENRIFFEKMGNGTEBGLKQNCVRAXI 363

Qy 353 DMQPNKSLRDPVYRRCNTDPHHRVGSKYVPTYPDFACPFVDALLEGYTHALRSSEYHNR 412

Db 364 DYKALNLTLRDPVYRRCNLPHHRTGSTWKIYPTYPDCVPVDAIEGYTHALRIEYDR 423

Qy 413 NAOYRRLIDMGARVEIYEPSRLNMYTLLSKRKLIMFVONKVEWTDPRPTVOGIV 472

Db 424 NAOYDMMLQALRLKRVHIMDFARINPFTLLSKRKLQMMVNDKLVGMDDRFPVGRVR 483

Qy 473 RGLKVEALLQFIQOGASKNLNMEMDKLMTINKKIIDPYCARTAVLKNQRYIFTLTN 532

Db 484 RRGMTVEGLRNFVLSQSPSRVINLNMNLIAFNAKVIIDPIAPRHATAVNFVKIHLBGE 543

Qy 533 GPEEPFVRLIRHKKFBCAGKATTFANRIWLDYADAANKGSEVTLMDGNALVKEIK 592

Db 544 APQEPKLEMKPKHKNNAVGEKVIYKYDVAVDKDDADVINVDEVTLMGKNALYTK-K 602

Qy 593 VESGVITELVGEILBESVTKTKLTWLDADIELVPLSIVEFDYLSKKKLEDEDFLD 652

Db 603 NDDG--SMVAKNLBEGFKTKKLTWLDATKQVVVDFVLDVDFHLITKRLIEDESPED 659

Qy 653 NLNPTRRREIPALGDANMNIRKGEIIOLEKGYRCDAPFIRSSKPVULFAIPDR 709

Db 660 FLTPQTERHMDALNDLVNKKMKIDIIQFEKGYRRLDA-LPKQKPFVFFTIIDPGK 715

RESULT 3

US-10-032-585-7526

Sequence 7526, Application US/10032585

Publication No. US20030180953A1

GENERAL INFORMATION:

APPLICANT: Terry, Roemer D.

APPLICANT: Bo, Jiang

APPLICANT: Charles, Boone

APPLICANT: Howard, Bussey

TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery

FILE REFERENCE: 10182-005-999

CURRENT FILING DATE: 2001-12-20

CURRENT APPLICATION NUMBER: US/10/032,585

NUMBER OF SEQ ID NOS: 8000

SOFTWARE: PatentIn version 3.1

SEQ ID NO 7526

LENGTH: 725

TYPE: PRT

ORGANISM: Candida albicans

US-10-032-585-7526

Query Match 41.2% Score 1537; DB 12; Length 725;

Best Local Similarity 47.3% Pred. No. 1.4e-135; Indels 22; Gaps 9;

Matches 303; Conservative 121; Mismatches 195;

Qy 81 EMLEAYAPTLGSGSEFENACL--FVDGFLASRTFLVGHGLTITADIAVNSNL--AGIQOR 134

Db 86 EMIKFALTUKLYVKNFKELAVLDKLDQHLNRSPIVGYSLADIAIGVLRANLMD-- 143

Qy 135 WESLRSKSKYONLVWENSI--DSEYKALNEVVAAPVGRKGIGKS--PAPSLKRYHD 189

Db 144 --SVYKNGVANIISRWVYLLADDKRFEGSVELMTKSLAEVRAKAAKSAKTAAGAAGKEAH- 200

Qy 190 SKDPSAPREVLDLPGAKVGVYCRFAPBEGGYLIGHAKAALLNKYFAERYOGRILYRFPDT 249

```

Db 201 ---KASFEIDLPGAMGVTRFRPEPSGYLTHGAKAVALNEYPANFKGLIIRFDDT 257
Oy 250 NPSKSNFEVENLKDIEITLGIKYDAVYTSDFPKLMEASLTKOKAYIDDPKQOM 309
Db 258 NPTKEKEFODSIIEEDLGLGKQDITSSDYFOTMDLAVKMKDKDNAACDDTPVDTM 317
Oy 310 RKEKMDIESRCRNNTVEENLSLW-KENVNGTERGMOCVCKLMDMODPNKSLRDPVYR 368
Db 318 REGVAVGASARRESVVENLRIFTTEEMKNGTEBGLKNCILAKIDYKAPNKALRPVYR 377
Oy 369 CNTDPHHRVSKKYKYPFYDPACPFVDALLEGVTHALRSSEYHNRNAQYRILQDMGLRRV 428
Db 378 CNLTPHRTGTWEMKYPFYDPCVPVDSIEGVTHALRTNEYRDRNPQYEMMQKTLGLRPV 437
Oy 429 EIVFESRLNMYTTLSSKRLMFPQNKVEBMDTRPFTVOGIVRGLKVALLOFILOQ 488
Db 438 AIMPGRNPFRTLLSKRKLQMPVDKNVSNWDDPRFPTVGRVRRGMTVGLRNFITISQ 497
Oy 489 GASKNLMEWDKMTINKKIIDPVCARHTAVLKQRYIFLTJNGPEEPFVRIIPRHKKF 548
Db 498 GPSNIIINLEWSVITAMNKIIDPAPRFTAVDAGNVSVKLNQPKPTYESKPKHKN 557
Oy 549 EGACKKATTFANRIMLDYADAAAINKGEVTLMDMGNAIYKEIKVESGITELVGEHLLE 608
Db 558 PEVGNKDVIFADQVLIQDQD-ADLTGEGEVTFMWGNIIIVSKNKEGDIIVKSEANHLLE 616
Oy 609 GSVTTKTKITWADIEELVPLSLVEFPYLSKKLLEDEDFDLNMLNCTREIIPALQDA 668
Db 617 GDFPKTSKRTITWLVYTKDKVEIDMDFDLITTKDLNDNDFEPTETBTEFHTKGFADL 676
Oy 669 NMRNIKGEIIOLEKGYRCDAPIRSSKPVLPFAIPDGR 709
Db 677 NVGKLKAGDIIQFERKGFIRVDKP-SEEGKPAVLTYTIDGK 716

```

RESULT 4

```

US-10-128-714-3375
; Sequence 3375, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Mengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3375
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3375

```

Query Match 39.4%; Score 1471.5; DB 15; Length 715;
 Best Local Similarity 43.0%; Pred. No. 26-129;
 Matches 310; Conservative 128; Mismatches 234; Indels 49; Gaps 14;

```

Oy 4 ALSEKSPPIISIIICAALVGLPLTINSLAAGSAPTLQF--ASGESLHGVNPILLYIAR 61
Db 26 SINARBPSPVISTYIEDAV-----LAEQKAVQYIGASGNPFLGLINAVQELRK 76
Oy 62 GASIASLSGRND-IEFGHVEWLEVAPTF--LSGSSEFNACLFPVDFLASFTFLVGHUT 118
Db 77 --DEPFLNSKQEKLE---NEWLSQLEAFAPLDKALDPKLRDLTHLLLSFFVGVALS 130
Oy 119 IADIAVMSNLAGIQRMESLRKSKKYQNLVWPFMSID-----SEYKALNEVVAAPFGK 172
Db 131 TADIALMGATRGNVAAVAALIKG-SLVVNTWTFYFLBLCMAWSTLEVLNQAVR----- 184
Oy 173 RGIGKSPASLKEKHVSDSPSAPEVDLPKAGVGVCRFAPSPSGYLIHGAKALANK 232
Db 185 -----EKKAARAKGASVDIALNLTEKG-VTRFPPEPSGYLIHGAKALAND 232
Oy 233 YFA-ERYOGRILVFPDDTNPSEKSENEFVENLKDIEITLGIKYDAVYTSDFPKLMEANR 291
Db 233 YFAHEKXNGTLVAFDDTNPSEKLEFQDAIIEEDLALMGIRPKDKSYTSDFDELQYAL 292
Oy 292 SLIQGKAYIDTFRQMKRKMGIISRCRNNTVEENLSLWKENVNGTERGMOCVCKG 351
Db 293 QIIXDGAAYADDTKEVVAEQRMNGKPSKRDAASVEENLARFBEKKGTPEGLRWCIRAK 352
Oy 352 LDMODPNKSLADPYVYRCNTDPHHRVSKKYVPFYDPACPFVDALLEGVTHALRSSEYH 411
Db 353 MSVNPKNKMDPYIYRCNPAHPHRTGTMKIYFYDPACIYVDSIEGVTHALRTIEYRD 412
Oy 412 RNAQYRILQDMGLRVEIYEFSSRLNMYTTLSSKRLMFPQNKVEBMDTRPFTVOGI 471
Db 413 RNPQYQWFLDTLKLKRNHQIMQFAMNPIRITLSSKRLTKLVNQGVMWDDPRPFTIRGI 472
Oy 472 VRGKLYEALIQFLQOGASKNLMEWDKMTINKKIIDPVCARHTAVLKQRYIFLT 531
Db 473 RRRGMTIPALREFLIKQSPSKNITNLDWTLTWTATNKKYIDVAPRHTAILKDKWKAIVK 532
Oy 532 NGPEEPFRIIPRHKKEGAKKATTFANRIMLDYADAAAINKGEVTLMDMGNAIYKEI 591
Db 533 GGPATPYTEKPKGKGPANVMKRVFQNTVYIQDQDKSKQDEBITLMSGNATYAKI 592
Oy 592 KVE--SGVITELVGEHLLEGSVKTKTKITWLA-DIEELVPLSLVEFPYLSKKLLEDE 648
Db 593 ETDPTSGIVAELELHLEHGFKTEKKVYTWLSREGDLIVELVDPYLNKOTLQDQD 652
Oy 649 DFLDNLPCTREIIPALGDANMRNIKGEIIOLEKGYRCDAPIRSSKPVLPFAIPDGR 708
Db 653 VLEEDVLMNTEFFRBDAAVADCNVAELKEGDIIOFERKGYRVDRAYV-PGKPAVLFIPIPTG 711
Oy 709 R 709
Db 712 K 712

```

RESULT 5

```

US-10-128-714-8375
; Sequence 8375, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Mengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066

```

Db 712 K 712

Db 593 NIPSGK 598

RESULT 7

US-10-369-493-4230
 ; Sequence 4230, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 4230
 ; LENGTH: 618
 ; TYPE: PRT
 ; ORGANISM: Neurospora crassa
 ; US-10-369-493-4230

Query Match 36.1%; Score 1349; DB 12; Length 618;
 Best Local Similarity 46.0%; Pred. No. 5.8e-118;
 Matches 285; Conservative 100; Mismatches 210; Indels 24; Gaps 10;

QY 102 VDGGLASTPLVGLTADIAVWSNLGIGQWESLRKSKYQNLVWFNSIDSEYKEA 161
 DB 12 LDKHLIATYDGLTGIDITKIMWLANG-NRAAVSFTRKSLANLAWFYIEENHEI 70
 QY 162 LNEVVAFAVGRGIGKSPAPSLKEKV-HDSKSPASPEVDLPGAKGVKVCVFAPEPSGYL 220
 DB 71 QAEI-----KADDAAKAKVAASAKGASYALALQNDQG-VTFLEPSPGYL 118
 QY 221 HIGHAKALLNKYFA-ERYQRLIVRPDITNPSKESNEFVENLKDITIGIKYDAVYTT 279
 DB 119 HIGHAKALLSDYFAHQYKQGLRLRLDITNPSKEQYDAIIEDLALMGIKDVTYTT 178
 QY 280 SDYPPKLEMAESLIGKQKAYIDTPKEQMKERMDIGESCRNNTVEENSLM-KEMVN 338
 DB 179 SDYDYIYDMCVMIKEGHAAYADDTDQDTRDQKWKGLASARRRSVEENRIRITEEKN 238
 QY 339 GTEGMOCVVRGKDMODPNKSLRDPVYYRCNTD-PHHRVSKYKVPYTYDFACPFVDAL 397
 DB 239 GTEGLKAKCIKAKLSVDNPNALDPIYRCNIEIPHRITGKMKMTFMTDFACPVDSH 298
 QY 398 EGVTHALRSSEYHNRNAYRILQDMGLRVEIYFSRLNMYTLLSKRKLMEVQNKV 457
 DB 299 EGVTHALRSTETDRNPQYQWFIIDTKLRQYVMDPAMNFIIRFLSKRLAKLVDSGV 358
 QY 458 EDWTDPPRPYQGIIVRGKLYEALIOFLQOGASKNLNMEMDKMTINKKIIDPVCAKH 517
 DB 359 WGMDDPMPRTIRGVRRRGMTIPALRDFILKOGPSRNVVMTDMTFWASNKEIDPIAPRH 418
 QY 518 TAVLKDGQ-RYIFLTLNGPEEPFVRILIPRHKKEGAGKATTPANRIMLDVADAANKG 575
 DB 419 TAITKCAVAVKAVGADAPABPVQOEKPKPKNKEVGKQVTFANELLMDADAKSPKDG 478
 QY 576 EEVTLMDGNAIYKIVESGVITELVGEHLLESGVTKTKITWLADI-ELVPLSLIVE 634
 DB 479 EEITIMGMNGFVNIDTSAAVIPAFAIEDNLADGVSTEEKVYWLMSKGTIVPAELMD 538
 QY 635 FDYILSKKLEDEDFLDNLNPTCRREIPALGDANMRNIRKGEIIOLEKRGYVRCDAPE- 693
 DB 539 FDYILITDVOLEBNMEDFLNPVETMEAWCDEAASLKKDIIIOLERRRGYVVDKGLN 598
 QY 694 ---IRSSKPVYLFAIPGR 709
 DB 599 DMKEGEEKVLFCIPGK 617

RESULT 8

US-10-369-493-5211
 ; Sequence 5211, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 5211
 ; LENGTH: 1149
 ; TYPE: PRT
 ; ORGANISM: Caenorhabditis elegans
 ; US-10-369-493-5211

Query Match 35.9%; Score 1339; DB 12; Length 1149;
 Best Local Similarity 39.5%; Pred. No. 1.4e-116;
 Matches 292; Conservative 132; Mismatches 243; Indels 72; Gaps 13;

QY 2 EALSFKSDPPISITCAKLVGLPLTINSLAAGSAPLQFA-----SGESLHGV 52
 DB 6 ELVVKARBEQPPVYSIALAASGFSL-----EKYQFSKQOLALNLDDELISND 55
 QY 53 NPILTYARAGSIA-SLSGNDIEFGHVEWLEAYP-----TFLSGSPFENACLPVD 103
 DB 56 VEIRITIAQSTDADSLSSIIDFAVVDITNFIAAVTKNDYSILGKDP----- 108
 QY 104 GFLASRTPLVGLTADIAVWSNLGIGQWESLRKSKYQNLVWFNSIDSEYKEALN 163
 DB 109 -----TKLIIDNSLTVDPAIFS-----VAHNNPQLK-----AKSGIIDKYLKKEPTL 150
 QY 164 EVAAPFGRKIGKSP--APSLKEKYHDSKDSAPAVDLPKAGKGVKVCVFAPEPSGYL 220
 DB 151 AAANFVGLYKSAAPATATASGKEK---KQEGKVELEPGAEKGVVAVFPPEASGYL 206
 QY 221 HIGHAKALLNKYFAEYQRLIVRPDITNPSKESNEFVENLKDITIGIKYDAVYTT 280
 DB 207 HIGHAKALLNYYQAFEBQGLIRFPDITNPAKENAHFEHVIKEDLSMLNIVPRWTHSS 266
 QY 281 DYFPKLEMAESLIGKQKAYIDTPKEQMKERMDIGESCRNNTVEENSLM-KEMVNGT 340
 DB 267 DHFEMLLTMECKLKEGKALVDDTDTETMNEREQDSDNRNSTPPEKNQOLMEEMKGS 326
 QY 341 ERGMQCVVRGKDMODPNKSLRDPVYYRCNTDPPHHRVSKYKVPYTYDFACPFVDAL 400
 DB 327 PKGILTCVBRMKIDMKSNNGAMRDPITYRCRKEBEHVIRGLKYKVPYDFCFPIVDSVEG 386
 QY 401 THALRSSEYHNRNAYRILQDMGLRVEIYFSRLNMYTLLSKRKLMEVQNKYEDW 460
 DB 387 THALRTTEYHNRDQYFFICDALGLRPHIMEVARIAMTNTVMSKRKLTFVVEGHEGW 446
 QY 461 TDPEFPVQGIIVRGKLYEALIOFLQOGASKNLNMEMDKMTINKKIIDPVCAKH 520
 DB 447 DDRRLPFRVGRVGRGLTVEGKQPIVAGSGRSRYVMEMDKINAFKQVLDVPAERTAL 506
 QY 521 LKQDRI-FTLTNGPEEPFVRILIPRHKKEGAGKATTPANRIMLDVADAANKG 579
 DB 507 DSTSPVLSIRLTDISDTSNV-SLHPKNABISGDVHKGKLLLEQVDAALKEGIVT 565
 QY 580 LMDGNAIYKIVESGVITELVGEHLLESGVTKTKITWLADI-----ELVPLSLIVE 635
 DB 566 FVVMGNKIKIGIEKGVAVITKISATIQDNTDYKTKTYVTLGIVKAEAKGTIPVADY 625
 QY 636 DYILSKKLEDEDFLDNLNPTCRREIPALGDANMRNIRKGEIIOLEKRGYVRCDAPIR 695

Db 626 DHIISKAIGDEEMKQFINPDSVHYTMGEPALIKVYKGDIIQIQKRGFYVDQPINP 685
Qy 696 SSK-----PVLFAPIDG 708
Db 686 KSELGSETPILLIAPIDG 704

RESULT 9
US-10-104-047-3100
; Sequence 3100, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104.047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3100
; LENGTH: 717
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3100

Query Match 34.5%; Score 1287.5; DB 12; Length 717;
Best Local Similarity 41.2%; Pred. No. 4.7e-112;
Matches 280; Conservative 127; Mismatches 244; Indels 29; Gaps 13;

Qy 46 GE-SLHGVNPIILIYARGASIASLSGKNDIEFGHVEMLEYAPFLSGSEFENACLFYDG 104
Db 43 GETTIKGEIDAICDYFISKCADQSLGCDAAE--TVKEWRHIASTTTKENVLEQ-IEKLDI 99
Qy 105 FLAARTFLVGHGLTADIAVMSNLAGIQRWESLRKSKYQNVLRWFMSIOSEYKEALNE 164
Db 100 YIARTYIVRVNFTIADISMAAI-----KLTVEIIPAKCTFASWYTYISA--IPSKQ 152
Qy 165 VVAAFVGRGIGSKPAPSLKEKVH-----DSKDPAPFVDPDYGAKVGVCFAPFAP 216
Db 153 TTGAVGSSLRMAAQASAPKKEIATKKEISLDGPREMPE--LPFAMGVVTRFPEBA 210
Qy 217 SGYLHIGAKAALINKYFAERYOGRILVRFDDTNSKESNEFVENLADITLGIKYDAV 276
Db 211 SGYMHIGHVAKAMLNYYAKRYOGKILRFDDTNSKEKEFEFTSIIRDIAKGIKADLF 270
Qy 277 TYTSDYFPKLMEMESLKGKAYIDTPEKQMRKERMDGIESRCRNTVEENLSLWKEM 336
Db 271 SHTSDYFPVILIDYARQMTREGIAFMQNDTQETMRKERMERKESKLRNTSPENLRLEBAL 330
Qy 337 VNGTERGQCCVRGLDMQDQPNKSLRDVYVRCNTDPHHRVGSKKYVPTYDFACPFVDA 396
Db 331 CRGEPEVODYCLRAKIDMKSDNGTIRDPVLYRFPYLTMTGDKKAYPCIDLACPIYDS 390
Qy 397 LEGVTHALRSSEYHDNAQYRIIQQDMGLRVEIYEFSLNMVNTLLSKRLMFWQKK 456
Db 391 IEGLYTHARTTEYKDRDQYMWIIOKALRLRPVHLVEFARLNQYTLMSKRLTMMVBEHE 450
Qy 457 VEDWTDPEFPYVQGLVRGKLYEALIOFTLQOGASKNLNMEWDLMTINKKIIDPVCA 516
Db 451 VDGDNDPFPYVQGLVRGKLYEALIOFTLQOGASKNLNMEWDLMTINKKIIDPVCA 510
Qy 517 HTAVLKQDQVIFTLNPGPEEPFRILPRHKKPEGAGKATTPANRIMDYADAALINKGE 576
Db 511 FMGVAKSPFVTLTLN-VQPGFIK-V-PNHPKDESQVSDIAGPVMLEFRDQGLKGE 568
Qy 577 EVTLMDGNAIVKEIKVES-GVITELVGBELHLEGSVTKTKITWLADIIBELVPLSLVEF 635
Db 569 EFTLMRWSAKVTIARADAGVAVTGTEPVPNGDFKNT-AKVMNVAACPEHNEHTVVEY 627
Qy 636 DYLLISKKLLED---EDFLDNLNPTCTREIIPALGDANNRNIKGGIIOLEKGYRCAP 692

Db 628 DDLTKTLEBTDNFEFLTSKDHPTKMEYTVIYNACKGLKENQIVQLEBGRGYRCDP 687
Qy 693 FIRSKPVLFAIPDGRQA 712
Db 688 DTENS-PAKLILIPDGKKA 706

RESULT 10
US-10-369-493-751
; Sequence 751, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369.493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 751
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-369-493-751

Query Match 20.2%; Score 753; DB 12; Length 554;
Best Local Similarity 35.2%; Pred. No. 8.7e-62;
Matches 189; Conservative 91; Mismatches 205; Indels 52; Gaps 14;
Qy 199 DLPAKYGKVCVRAPEPSGLIHGAKAALINKYFAERYOGRILVRFDDTNSKESNEF 258
Db 19 DLASGKTTVHTRPPEPNGLIHGAKSICLNGLIADYKQCNLRFDDTNPVKEDIEY 78
Qy 259 VENLKDIEITLGIKYDA-VTYTSDYFPKLMEMESLKGKAYIDTPEKQMRKER--- 313
Db 79 VESTIKNDVENWGFHMSGNVRRSSDYFDQLAHAYIELLNKGLAYVDELPEQIRREYRGLT 138
Qy 314 MDGIESCRNNTVEENLSLWKEM-VNGTERGQCCVRGLDMQDQPNKSLRDVYVRCNTD 372
Db 139 QPGKNSPYRDRSVEENIALFEKMRAGGFEBG-KACILRAKIDMASPFYMRDPVLYRIKFA 197
Qy 373 PNHVGSKKYKPYDYDACPFDVALBGVTHALRSSEYHDNAQYRIIQQDMGLR-RVEIY 431
Db 198 EHHQTNKWCICYPMYDTHCISDLBGLTHSLCTLEFODNRRLYDWDNDITTFVHPRQY 257
Qy 432 EFSRLNMYTLLSKRLMFWQNKVEDWTDPRFPYVQGLVRGKLYEALIOFTLQOGAS 491
Db 258 EFSRLNLEYTMSRKRLNLVYTDKGVGWDPRMPTISGLRRGYTAASIREFCRIGVT 317
Qy 492 KNLNLMWDLMTINKKIIDPVCAHTAVLKQDQVIFTLNPGPEEPFRILPRHKKPEGA 551
Db 318 KQDWTIEMASIESCIRDELNENAPRAMAVIDPVGLVIENYQGBE--MVTMPPNPKPEM 375
Qy 552 GKATTPANRIMDYAD-AAAINK-----GEVTLMDGNA-IVKEIKYESGVITELV 602
Db 376 GSRVPPSGELWIDRDFREANQYRLVLGKEVRLR--NAVYIAERVEKA----- 427
Qy 603 GELHLEGSVKT-----TKLK--ITWLADIIBELVPLSLVEFDYLSKKK 643
Db 428 -----EGNITITICTYADLTISKDPADGRKKYGVIIHVSAAHAL-PVEIRLYDLFSVPN 481
Qy 644 LEEDEDFLDNLNPTCTREIIPALGDANNRNIKGGIIOLEKGYRCAPFIRSSKPV 700
Db 482 PGAADDFLSVINPESLVIKQGFAPSPSKDAVAGAKAFQFEREGYCLDSRHSHTAEKPV 538

RESULT 11


```

; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13759
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
US-10-369-493-13759

Query Match      18.9%; Score 704.5; DB 12; Length 555;
Best Local Similarity 34.8%; Pred. No. 3.3e-57;
Matches 189; Conservative 87; Mismatches 206; Indels 61; Gaps 18;

QY 197 EVDPGAKGVKVCVRFAPEPSGYLHIGAKAALLNKYFAERYOGRILVFPDTPNPKESKN 256
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 17 QADDSGKHTQIVRFRPEPNGLIHGAKSICVNFGLAQEGVTHLRFDDTPNPKEDQ 76
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 257 EFVENLKDIETLGIKYDA-VTYSDFPKLMEASLIIKOGKAYIDTPKEQMKER- 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 77 EYDIAESDVKMLGFEWSEGEVRYASQYFDQLHDMAVELIKAGKAYVDDLTEBOAKEYGS 136
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 314 --MDGIESCRNNTVEENLSLMKEVNGTERGMQCCVVGKLDMDPNKSLRDPVYRCNT 371
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 137 LTERGKSPFRDRSVEENLDMFARRAGEFPDGAFLAKIDMASPNMLRDPVYRIRH 196
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 372 DPHRVGSKYKVPYTPDFACPFVDALSGVTHALRSSEYHDINAQYRILIDMGL-RVEI 430
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 197 AHHTQGTGKCIYRPNYDFTHQSDALRIGTISICLFEFSHRPLYEWLDMALPVARRQ 256
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 431 YFSRLNNVTLLSKRLKLMFVQNKKEVDWTDPRFTVQGIYVRGLKVEALLQFTLQOGA 490
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 257 YFSRLNNVTLLSKRLKLMFVQNKKEVDWTDPRFTVQGIYVRGLKVEALLQFTLQOGA 490
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 491 SKNLNMEMDKLMTINKKIIDPVCAHRTVLKDQVIFLTNGPEEPVRI-LPHKKFE 549
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 317 NRSQGVDFGMLFESIRDDLDANARAWCVLRPLV--ITNPPDQVENLELPHPKPE 374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 550 GAGKATTPANRIMLDVAD-----AAAIKGEVTLMDGNAIVK--EIKVESGVI 598
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 375 ELGRQRLPFABEYIYDRDPMEEPPKGYKLEPNCEVLR--GSVITADALIKADANI 432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 599 TEL-----VGLHLGSAVTKTKITWL--ADIELVPLSLVEEDYI--SKKL 644
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 433 VELRCSYDPTLGK-NPEG--RKVGVIHWVPAASICEVRL---YDRLFRSPNPKA 485
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 645 EEDDFLDNLNLP-----CTRREIPALGDAMNIRGELIQLERKGYRCDAPIRS 696
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 486 EDSASFLDNINPGLLQVLVTC--RAEPLGNAQPE---DRFQEREGYFVAD--TKD 535
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 697 SKP 699
   |||
DB 536 SKP 538

RESULT 14
US-10-369-493-412
; Sequence 412, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21

```

```

; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 412
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-369-493-412

Query Match      18.8%; Score 703; DB 12; Length 554;
Best Local Similarity 33.5%; Pred. No. 4.6e-57;
Matches 177; Conservative 94; Mismatches 204; Indels 54; Gaps 13;

QY 199 DLPKAKGVKVCVRFAPEPSGYLHIGAKAALLNKYFAERYOGRILVFPDTPNPKESNEF 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 19 DLATGKHTSVTRFRPEPNGLIHGAKSICVNFGLAQEGVTHLRFDDTPNPKEDIEY 78
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 259 VENLLKDIETLGIKYDA-VTYSDFPKLMEASLIIKOGKAYIDTPKEQMKER- 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 79 VESIKQDVQWLGFPQMSDIRSSDYFDQLHQAELINKGLAYVDELSPDIRRYRGTLK 138
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 314 MDGIESCRNNTVEENLSLMKEVNGTERGMQCCVVGKLDMDPNKSLRDPVYRCNTDP 373
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 139 EPGKNSPYRDRSVEENLALFEKMRAGEFAEGKACLRAKIDMASPFMWVRDPVLYRIKFAE 198
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 374 HHRVSKYKVPYTPDFACPFVDALSGVTHALRSSEYHDINAQYRILIDMGLR-RVEIYE 432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 199 HHQGNKRCIYPMYDFTHCISDALRIGTISICLFEQDNRRLYDVLNDITTDCHPROYE 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 433 FSRLLNNVTLLSKRLKLMFVQNKKEVDWTDPRFTVQGIYVRGLKVEALLQFTLQOGASK 492
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 259 FSRLLNNVTLLSKRLKLMFVQNKKEVDWTDPRFTVQGIYVRGLKVEALLQFTLQOGASK 492
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 493 NLNMEMDKLMTINKKIIDPVCAHRTVLKDQVIFLTNGPEEPVRI-LPHKKFE 552
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 319 QNNVEMAALSCIRDDLNEHAPRAMAVLDPVKLV--IENPBEVILLTMNHNKPEMG 376
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 553 KKATTPANRIMLDVAD-----AAAIKGEVTLMDGNA-IVKEIKVESGVIITELVG 603
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 377 SREVPFRELITDRAIDREANROYKRLVNGKEVRLR--NAVYKAEKEDKDA----- 427
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 604 ELHLEGSAVTK-----TKLK--ITWLADIELVPLSLVEEDYISKKKL 644
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 428 ----EGNITTYICHYDQTLNKDPADGRKVGVIHWVSAV-HAAPAEYVLDRLFSVNP 482
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 645 EEDDFLDNLNLPCTRREIPALG--DANNIRKGEIITQLERKGYRCDA 691
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 483 GAEDDFLSTLN--TESLVVRKGEVPGVLVNAMLEKTYQFEREGYFVADS 529
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
US-10-369-493-9393
; Sequence 9393, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9393
; LENGTH: 573
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-369-493-9393

Query Match      18.3%; Score 684.5; DB 12; Length 573;

```

Best Local Similarity 33.3%; Pred. No. 2.7e-55; Matches 179; Conservative 86; Mismatches 217; Indels 55; Gaps 10;

```

Oy 211 RFAPPSGYLHGHAKAALLNKYFAERYOGRLIVRPDDTNPSEKSENEFVENLKDIOETLG 270
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 32 RPPPPNGYLIHGHAKAICLDPGVAAREFGHCTLRMDTNPSEKEDPAFAAIQEDVSWIG 91
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 271 IKYDAVTVTSDFPKLMEWASLIKQKAYIDDTPEQWKKER---MDGIESRCRNMTV 326
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 92 FHMNALRHTSDYFEVLVLAAEKLIADGKAYVCDLNSQVREYRGTLTEAGRSPWRERSP 151
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 327 EENLSLMKEMVNGTERGQCCVVRGKLDMDPNKSLDPPVYTRCNTDPHHRVSKTKVYPT 386
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 152 DENLELFRQMRAGTPPDGTRTLRAKIDWASGNINLRDPALYRIKHVEHQNTGTWPIYPM 211
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 387 YDFACPFVDALRGVTHALRSSEYHDMNAQYRILDMGLRVEI----- 430
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 212 YDFALSDAIEGITHSLCTLEFED---HRPLYDMCINHVDLPNNSHLKPILLDKGFP 266
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 431 -----YESRLNMVYTLISKRLLMFVQNKKYEDWTPRPFTVQGIVRGKLKVEALI 482
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 267 QEPSPQRIEFSRLNINVTWMSKRKLTLVDEKLVEGMDPPMYTLQGLRRRGYTPPALMR 326
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 483 QPILQOGASKNLNLMEMDKLMTINKKIIDPVCARHTAVLKQGVIFTLTNGPEEPFVR-I 541
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 327 LEVERIGISKONSIDFSVLENCLRENLDITLAPRMATLAPMKLV--LTNLPEDHEQLI 384
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 542 LPRHKKFEGAGKKATTTPANRIMWLDYADAALN-----KGEVTLMDWGNALYKE-1K 592
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 385 FPNHFKDPTQGRTRVPFSRELTIERDDFSEVPKGMKRLVPGGEVRLRGAGIARIDEVVK 444
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 593 VESGVITELVGEIH-----LEGSVKTWKIKITWLADIIELVPLSLVEFDYLISKKLEE 646
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 445 NAEGHIIALHGWLDPTSRPGWGAHRXVKGTIHWVS-APHAVAAEIRLYDRLSIEKDD 503
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 647 DEB---FLDNINPCTREIIPALGDAMNENIKRGEIIOLEKRGYRCDAPIRSSKPV 700
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 504 NTDGKTYRDFLNPDSKRVVHGYLEPAAQTAPEHAFOFERLGVFVTDRIHDHATHPV 560
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Search completed: January 25, 2004, 17:38:18
Job time : 248 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2004, 16:58:19 / Search time 71 Seconds
(without alignments)
968.459 Million cell updates/sec

Title: US-09-831-683B-10

Perfect score: 3734

Sequence: 1 MEALSFKSDSPISIIICAA.....SSKPVVLFALPDGRQASLS 715

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2657	71.2	719	2	TS2043
2	2644.5	70.8	728	2	T01200
3	1574.5	42.2	716	2	T37830
4	1562	41.8	724	1	S53934
5	1498	40.1	1714	1	S18644
6	1437	38.5	1440	1	SYHUC7
7	1339	35.9	1149	2	T27567
8	757.5	20.3	555	2	AC0585
9	753	20.2	554	1	SYECOT
10	747	20.0	554	2	F90717
11	747	20.0	555	2	F85567
12	743.5	19.9	555	2	AB0321
13	721	19.3	556	2	H82254
14	710.5	19.0	557	2	G64118
15	704.5	18.9	556	2	F83421
16	690	18.5	571	2	A84978
17	683.5	18.3	580	2	G82693
18	666.5	17.8	562	2	B81069
19	664.5	17.8	573	2	B81069
20	653	17.5	571	2	G75165
21	636.5	17.0	794	2	T09643
22	624	16.7	570	2	F71049
23	594	15.9	809	1	SYBYOT
24	589.5	15.8	553	1	H64471
25	575	15.4	852	2	A75253
26	573	15.3	786	2	T26811
27	572	15.3	811	2	T40275
28	561.5	15.0	775	1	I37422
29	559	15.0	551	2	D69282

30	537.5	14.4	553	2	B69167	glutamate-tRNA lig
31	530	14.2	552	2	S65787	glutamate-tRNA lig
32	527.5	14.1	729	2	D86383	probable glutaminy
33	504	13.5	575	2	S75395	probable glutamine
34	477.5	12.8	544	2	A72459	probable glutaminy
35	474.5	12.7	586	2	PC4398	glutamine-tRNA lig
36	381.5	10.2	586	2	B84271	glutamine-tRNA lig
37	275.5	7.4	485	2	C97022	glutamyl-tRNA synth
38	271	7.3	473	2	D70405	glutamate-tRNA lig
39	266.5	7.1	468	2	S21172	glutamate-tRNA lig
40	255.5	6.8	464	2	F82019	glutamate-tRNA lig
41	255	6.8	469	2	G72264	glutamate-tRNA lig
42	255	6.8	484	2	S73490	glutamate-tRNA lig
43	250	6.7	463	2	A81337	glutamate-tRNA lig
44	247	6.6	487	2	F72200	glutamate-tRNA lig
45	245.5	6.6	467	2	G84937	glutamate-tRNA lig

ALIGNMENTS

RESULT 1

TS2043
probable glutamate-tRNA ligase (EC 6.1.1.17) [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Jun-2002

C:Accession: T52043

R:Day, I.S.; Golovkin, M.; Reddy, A.S.

Biochim. Biophys. Acta 1399, 219-24, 1998

A:Title: Cloning of the cDNA for glutamyl-tRNA synthetase from Arabidopsis thaliana.

A:Reference number: 224836; PMID:9765600; PMID:9765600

A:Accession: T52043

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-719 <DAY>

A:Cross-references: EMBL:AF067773; PIDN:AAC36469:1

C:Superfamily: Yeast glutamate-tRNA ligase; glutamine-tRNA ligase homology

C:Keywords: ligase

Query Match	71.2%	Score 2657	DB 2	Length 719
Query	Local Similarity	69.0%	Pred. No. 2.4e-169	Matches 491; Conservative 106; Mismatches 107; Indels 8; Gaps 5;
1	LSFSDSPISIIICAAATVGLPLTNHSLAAGSAPTIQFASGESLHCVNPIITLARGA- 63			
2	6 LSFSDSPISIIICAAATVGLPLTNHSLAAGSAPTIQFASGESLHCVNPIITLARGA- 65			
3	64 SIASLGGNDIEFGHVVEMLEVAFTPLSGSEFENACLFVDGFLASRTFLVGHGTTIADIA 123			
4	66 KLPDPYGNNAFPDSSQIDEMVDYASVFGSGSEFENACRVKYLESSFTLVGHSLSDVA 125			
5	124 VMSNLGIGQWESLRKSKYQNLVRWPNISIDSEYKALNEVAAFYGKRGIGS-PAPS 182			
6	126 IWSALAGTQWRWESLRKSKYQNLVRWPNISIDSEYKALNEVAAFYGKRGIGS-PAPS 184			
7	183 LKEKVN---DSKPSAPVDLPGAKGYCVAPAPPSGVIHTGHAKALNLYFARY 238			
8	185 SKSQQAQVADGQKRPVDLPEALIGKYLRAPEPSGLHTGHAKALNLYFARY 244			
9	239 QGRLLVFDDTNPEKSENEFENLKDIEITGICYDAVVTSDYFPKLMEAAELIRQK 298			
10	245 QGEYIVAFDDTNPAKSENEFENLKDIEITGICYDAVVTSDYFPKLMEAAELIRQK 304			
11	299 AYIDDTPEQWRKERMKGIESRCNNVTVEENLSLWKMVNGTERGMOCVVRGLMDWDPN 358			
12	305 AYVDDTPEQWRKERMKGIESRCNNVTVEENLSLWKMVNGTERGMOCVVRGLMDWDPN 364			
13	359 KSJADPYVYRCNDPDRHVRGSKYVYPTYPACPFVDALEGTATLSSEVHDNAOYR 418			
14	365 KAMDPIYVYRCNDPDRHVRGSKYVYPTYPACPFVDALEGTATLSSEVHDNAOYR 424			
15	419 ILDDMGLRVEIYFSLNNVYTLTKRKLMLFQNNKVEDWTDPRPPTVQGIYRGLKV 478			

Db 425 VLEDMGLQVQLYFERSRLNLTLSKRLLMFVQTGLVDGMDPRPFTVQGIYRGLKI 484
Qy 479 EALIQFIQOGASKNLNMEDKLTMTINKKIIDPVCARHTAVLKQDQVIFTLTNGPEEPF 538
Db 485 EALIQFIQOGASKNLNMEDKLTMTINKKIIDPVCARHTAVLKQDQVIFTLTNGPEEPF 544
Qy 539 VRIIPRHKKEGAGKATTPANRITLVDYADAANKGSEVTLMDGNMIVKEI-KVESGV 597
Db 545 VRMI PKHKKEGAGKATTPKTSIWLBEADASISVGEVTLMDGNMIVKEITKDEGR 604
Qy 598 ITEVLGELHSGVYKTKITLMLADIEELVPLSLVEEDYILSKKLEDEDFLDNLNCP 657
Db 605 VTLASGVNLQGSYKTKITLMLPDTMELVNLTLTEEDYILTKKLEDDDBVDVAFVNP 664
Qy 658 TRREIPALGDANMRNKRGEIITLERKGYRCDAPFIRSKPVLFAIPDG 709
Db 665 TKKETLALGDSNMRLKCGDVIQLERKGYFRCDVFPVSKSPVLFSPDG 716

RESULT 2

T01200
probable glutamate-cRNA ligase (EC 6.1.1.17) F21E10.12 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 03-Jun-2002
C:Accession: T01200
R:Davidson, S.; Kohlfing, T.; David, M.; O'Brian, D.
submitted to the EMBL Data Library, April 1998
A:Description: The sequence of A. thaliana F21E10.
A:Reference number: 214258
A:Accession: T01200
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-728 <DAV>
A:Cross-references: EMBL:AF058914; NID:G3047074; PID:G3047084; GSPDB:GN00063; ATSP:F21E1
A:Experimental source: cultivar Columbia
C:Genetics:
A:Gene: ATSP:F21E10.12
A:Map position: 5
A:Introns: 47/2; 89/3; 141/1; 503/3; 659/3
C:Superfamily: Yeast glutamate-cRNA ligase; glutamine-cRNA ligase homology
C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis
F:223-499/Domain: glutamine-cRNA ligase homology <EBL>

Query Match 70.8%; Score 2644.5; DB 2; Length 728;
Best Local Similarity 68.2%; Pred. No. 1.7e-168;
Matches 492; Conservative 105; Mismatches 107; Indels 17; Gaps 6;
Qy 5 LSPKSDSPRITICAKLVGLPLTNHSLAAGSAPTLQFAGSESLHGVNPIILYARGA- 63
Db 6 LSPKSDSPRITICAKLVGLPLTNHSLAAGSAPTLQFAGSESLHGVNPIILYARGA- 65
Qy 64 SIASLSGKNDIEFGHV-----EWLEVAPTFLSGSEFENACLPYDGFILASRTFLVG 114
Db 66 KLDPDYGNNAAPDSQVSLICIMKIDEMVDYASVSSGSEFENACGRDVKYESTFLVG 125
Qy 115 HGLTIADIAVMSNLGICQWRWESLRKSKYQNLVWPNISIDSEYKALNEVVAALFVGRG 174
Db 126 HSLTIADIAVMSNLGICQWRWESLRKSKYQNLVWPNISIDSEYKALNEVVAALFVGRG 184
Qy 175 IGKS-PASLSKEKVA-----DSKDPAPAEVDLPAGAVGKVCVFAPEPSGYLHIGAKAL 229
Db 185 SGKPVAAKSKDSQAVKGDQDKGKPEVDLPKAEVDFGKLPFAPEPSGYLHIGAKAL 244
Qy 230 LNKYFAERYOGLIYRFDTPNPKSNEFENLIDQIETLGIKYDAVYTSDFPKLMMEM 289
Db 245 LNKYFAERYOGLIYRFDTPNPKSNEFENLIDQIETLGIKYDAVYTSDFPKLMMEM 304
Qy 290 AESLIKQKAYIDTPEQMRKEMDGLIESCRNNTVEENSLWKEMVNGTERGKQCCVR 349
Db 305 ABLKMRBEKAYVDTPREQMRKEMDGLIESCRNNTVEENSLWKEMVNGTERGKQCCVR 364
Qy 350 GKLMDQDNKSLRDVYVYRCNDPHHRVGSKYKVPYTFDAPCPVDALBEGVTHALRSSEY 409

Db 365 GKEMQDPNKAMRDVYVYRCNDPHHRVGSKYKVPYTFDAPCPVDALBEGVTHALRSSEY 424
Qy 410 HDNNAQYRILIQDGLRVEIYERSRLNMQYTLTSLKRLLMFVQNKYVEDMTDRFPVQ 469
Db 425 HDNNAQYRILIQDGLRVEIYERSRLNMQYTLTSLKRLLMFVQNKYVEDMTDRFPVQ 484
Qy 470 GIVRGLKVEALIQFIQOGASKNLNMEDKLTMTINKKIIDPVCARHTAVLKQDQVIF 529
Db 485 GIVRGLKVEALIQFIQOGASKNLNMEDKLTMTINKKIIDPVCARHTAVLKQDQVIF 544
Qy 530 LTNPEEPFVRIIPRHKKEGAGKATTPANRITLVDYADAANKGSEVTLMDGNMIVK 589
Db 545 LTPDEPDEPVRMI PKHKKEGAGKATTPKTSIWLBEADASISVGEVTLMDGNMIVK 604
Qy 590 EI-KVESGVITELVGEILBGSVYKTKITLMLADIEELVPLSLVEEDYILSKKLEDE 648
Db 605 EITKDEBERVTLASGVNLQGSYKTKITLMLPDTMELVNLTLTEEDYILTKKLEDD 664
Qy 649 DFLDNLNCPTRREIPALGDANMRNKRGEIITLERKGYRCDAPFIRSKPVLFAIPDG 708
Db 665 EVADFVNMTKETLALGDSNMRLKCGDVIQLERKGYFRCDVFPVSKSPVLFSPDG 724
Qy 709 R 709
Db 725 R 725

RESULT 3

T37830
probable glutamate-cRNA ligase (EC 6.1.1.17) - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Jun-2002
C:Accession: T37830
R:Devlin, K.; Churcher, C.M.; Bartell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1997
A:Reference number: 221737
A:Accession: T37830
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-716 <DEV>
A:Cross-references: EMBL:Z96849; PID:CA11515.1; GSPDB:GN00066; SPDB:SPAC17A5.15c
A:Experimental source: strain 972h-; cosmid c17A5
C:Genetics:
A:Gene: SPDB:SPAC17A5.15c
A:Map position: 1
C:Superfamily: Yeast glutamate-cRNA ligase; glutamine-cRNA ligase homology
C:Keywords: ligase

Query Match 42.2%; Score 1574.5; DB 2; Length 716;
Best Local Similarity 49.5%; Pred. No. 4.3e-97;
Matches 319; Conservative 102; Mismatches 201; Indels 23; Gaps 10;
Qy 70 GKNIDIEFGHVEMLEVAPTFLSGS-----EFENACLPYDGFILASRTFLVGHGLTIADIAV 125
Db 81 GENDRSL--VESWYETA-SALAGNHNFLLESLLAQDDHLIMSLFVGYSLTSDPSIW 137
Qy 126 SNLAGICQWRWESLRKSKYQNLVWPNISIDSEYKALNEVVAALFVGRGIGKSPAPSLKE 185
Db 138 GALKSNMMAAGAVR-TQGYFNLAIRWYKFMDSQ--NAVSVMTEEFKAVNIRK-----KQ 188
Qy 186 KYHDSKDSAPAEVDLPKAKVGVCFAPAPESGYLHIGAKALLNKVFARYOGLIVR 245
Db 189 K--SSGFNY-EIGLPDIDKVVTRPPESGYLHIGAKALNQFAFKYHGLKIVR 244
Qy 246 FDDTNPSEKSNFENLIDQIETLGIKYDAVYTSDFPKLMMEMAESLIKQKAYIDTDP 305
Db 245 FDDTNPSEKSNFENLIDQIETLGIKYDAVYTSDFPKLMMEMAESLIKQKAYIDTDP 304
Qy 306 KEQMRKEMDGLIESCRNNTVEENSLWKEMVNGTERGKQCCVRGKLDMDPNKSLRDV 365
Db 305 VETMRHERTEGIPSKHRRPIEESLEILSEMDKSDVGLKNCIRAKISYENPNKAMRDPV 364
Qy 366 YTRCTDPHHNVGSKYKVPYTFDAPCPVDALBEGVTHALRSSEYHDNNAQYRILIQDGL 425

C:Accession: S18644
 R:Cerini, C.; Kerjan, P.; Aebler, M.; Gratecos, D.; Mirande, M.; Semeriva, M.
 EMBL J. 10, 4267-4277, 1991
 A:Title: A component of the multisynthetase complex is a multifunctional aminoacyl-tRNA
 A:Reference number: S18644; MUID:92097547; PMID:1756734
 A:Accession: S18644
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1714 <CER>
 C:Cross-references: GB:M74104; NID:G157563; PIDN:AAA28594.1; PID:G157564
 C:Genetics:
 A:Gene: FlyBase:Act5-01upro
 A:Cross-references: FlyBase:FBgn0005674
 C:Superfamily: Drosophila multifunctional amino acid-tRNA ligase; amino acid-tRNA ligase
 C:Keywords: aminoacyl-tRNA synthetase; ATP; ligase; multifunctional enzyme; protein bios
 F:204-480/Domain: glutamine-tRNA ligase homology <EGL>
 F:755-800/Domain: amino acid-tRNA ligase repeat homology <ATL1>
 F:827-872/Domain: amino acid-tRNA ligase repeat homology <ATL2>
 F:901-946/Domain: amino acid-tRNA ligase repeat homology <ATL3>
 F:980-1025/Domain: amino acid-tRNA ligase repeat homology <ATL4>
 F:1055-1100/Domain: amino acid-tRNA ligase repeat homology <ATL5>
 F:1129-1173/Domain: amino acid-tRNA ligase repeat homology <ATL6>

Query Match 40.1%; Score 1498; DB 1; Length 1714;
 Best Local Similarity 44.2%; Pred. No. 2,1e-91;
 Matches 321; Conservative 126; Mismatches 247; Indels 32; Gaps 13;

1 MEAALSFESKDSPPISIIICAKLVG--LPITINHSIAASAPITOPASGSL--HGVRNII 56
 1 MSIKKALNNPISGLTAHLINTIVEVEIWSKEETS---LQPPDRRLVCHSNNVL 57
 57 LYIARGASIASLGNKDIEFGHVEWLEAPFLSSEFENACLFVDFGLASRTFLVGHG 116
 58 PALAAADYLYGRTAERTQIDHMLSFSLT--CEDDISWALSFITSPLEPPVTVLVANK 115
 117 LTIDIANVSNLAGICGEMESLRKSKYONLVRMNSIDSEKELNVVAFAVGKRGIG 176
 116 LTIDFALFNEM--HSRYEFALAKGIDPOHVQWYDLITTAQ--PLIQVLOSLEPDAYK 170
 177 KSPAPSLKE--KVHDSKOPSAPEVDLPQAKVGCYAFAPSPSGYLIGHAKALANKY 233
 171 RSPQSSKEQTAKTGERGQ--EKGFDLPGAEKGVAFPPAPSGYLIGHAKALANKY 229
 234 FAERYGSLIVRFDPTNPKSNEFVNLKDIEFLGKYDAVYTSDFPKLMEASL 293
 230 YALVQGLTLMRFDPTNPAKETVEENYVIGLQIKPDPVFTHSNVFDMLDYCVRL 289
 294 IKQGAIVYDTPPKEMKERNKDIGSRNNVNEISLMEKMNVTGRCMOCCVRGLD 353
 290 IKESKAYVDDTPPEKMLEREQVRVSNRNSVEKNLSLMEEMVSGSEKQNTACAKAID 349
 354 MODPKSLRDPVYVYRCNTDPHHRVGSKYKVPYTFACPFVDALRGVTHALRSSEYHRN 413
 350 MSSPNCGRDPTIYRCNKEPHRRTTKYKVPYTFACPIYDALINVTHTLTTHYHRD 409
 414 AQYRIIDMGRLRVEIYEFSLNNVYTLSSKRLMVFVQNKVEDWTDPRFPVQIVR 473
 410 DQFYFIDALKLRKYIWSYSLNNTNTVLSKRLTWFDVSGLVGMDPRFPVQIVR 469
 474 RGLKVEALIOPILOGASKNLMEWDKLTNTKTIIPVCARHNAVLDQGVFTTLNG 533
 470 RGMTVEGKEFLIAGSSKSVFNNWDMKIWANKVIDIARVYALKEKRVINAVGA 529
 534 PEEFVRV--LPRHKKFEGAGKATTFANRIMLDVADAANKGEVTLMDGNALVKRI- 591
 530 KVE--RQVSVHPDESLGKKTUULGRITIDYDAELKGEVATITMNNILIKYVN 586
 592 KVESGIVTELVELLSESVKTKTKITWLA---DIEELVPLSLVEFYLLSKKLEBDE 648
 587 KDAENITSVDAALNENKDKFKTKLTLTWLAVEDPSAVPFPFCYFNITISKAVLGDE 646
 649 DFLDNLNCTREIRIYALGDAMNRNKRGEIITLEKGYRCAPIRSS-----KPYVL 702

Db 647 DFKQFIGHTRDEVMVLGDELKCKKGGDIIQLQRGPFKVDVAVLPPSGVTNVPVL 706
 Qy 703 FAIPDG 708
 Db 707 FSIPDG 712

RESULT 6
 SYNDUT
 multifunctional aminoacyl-tRNA synthetase - human
 N:Alternate names: glutroyl-tRNA synthetase; glutaminyl-tRNA synthetase
 N:Contains: glutamate-tRNA ligase (EC 6.1.1.17); proline-tRNA ligase (EC 6.1.1.15)
 C:Species: Homo sapiens (man)
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jul-2002
 C:Accession: A38663; S03424; S00969; S38809
 R:Felt, R.; Knippers, R.
 J. Biol. Chem. 266, 1448-1455, 1991
 A:Title: The primary structure of human glutaminyl-tRNA synthetase. A highly conserved c
 A:Reference number: A38663; MUID:91107633; PMID:1988429
 A:Accession: A38663
 A:Molecule type: mRNA
 A:Residues: 1-1440 <RET>
 A:Cross-references: GB:X54326; NID:G31957; PIDN:CAA38224.1; PID:G31958
 A:Note: the cited Genbank accession number, X54327, is not in release 101.0
 R:Knippers, R.
 submitted to the EMBL Data Library, April 1988
 A:Reference number: S03424
 A:Accession: S03424
 A:Molecule type: mRNA
 A:Residues: 96-382, 'T', 384-497, 'IGATSTLQ', 506, 'YT', 509, 'WOME', 514, 'SYL', 518, 'MOSLIWKT', 5.
 A:Cross-references: EMBL:X07466; NID:931769; PIDN:CAA30354.1; PID:9825664
 R:Thoenes, P.; Felt, R.; Schrey, B.; Kunze, N.; Knippers, R.
 Nucleic Acids Res. 16, 5391-5406, 1988
 A:Title: The core region of human glutaminyl-tRNA synthetase homologues with the Escheri.
 A:Reference number: S00969; MUID:88262551; PMID:3290852
 A:Accession: S00969
 A:Status: nucleic acid sequence not shown.
 A:Molecule type: mRNA
 A:Residues: 96-146, 'T', 148-191, 'T', 193-382, 'L', 384-416, 'G', 418-497, 'IGATSTLQ', 506, 'YT', 5.
 HO>
 A:Cross-references: EMBL:X07466
 R:Cerini, C.; Kerjan, P.; Aebler, M.; Gratecos, D.; Mirande, M.; Semeriva, M.
 EMBL J. 10, 4267-4277, 1991
 A:Title: A component of the multisynthetase complex is a multifunctional aminoacyl-tRNA
 A:Reference number: S18644; MUID:92097547; PMID:1756734
 A:Contents: annotation; demonstration of glutamyl- and prolyl- tRNA synthetase activities
 R:Kaiser, B.; Eberhard, D.; Knippers, R.
 J. Mol. Evol. 34, 45-53, 1992
 A:Title: Exons encoding the highly conserved part of human glutaminyl-tRNA synthetase.
 A:Reference number: S38809; MUID:92211721; PMID:1556743
 A:Accession: S38809
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 96-382, 'L', 384-463 <KAT>
 C:Genetics:
 A:Gene: GDB:EPBS; QPRS; QARS
 A:Cross-references: GDB:126609; OMIM:138295
 A:Map position: 1432-1442
 C:Superfamily: human multifunctional amino acid-tRNA ligase; amino acid-tRNA ligase repe
 C:Keywords: aminoacyl-tRNA synthetase; ATP; duplication; ligase; protein biosynthesis
 F:125-485/Domain: tRNA-charging <GLN>
 F:126-402/Domain: glutamine-tRNA ligase homology <EGL>
 F:126-402/Domain: glutamine-tRNA ligase repeat homology <ATL1>
 F:677-733/Region: 57-residue repeat
 F:688-733/Domain: amino acid-tRNA ligase repeat homology <ATL2>
 F:750-806/Region: 57-residue repeat
 F:761-806/Domain: amino acid-tRNA ligase repeat homology <ATL3>
 F:828-884/Region: 57-residue repeat
 F:833-884/Domain: amino acid-tRNA ligase repeat homology <ATL3>

Query Match 38.5%; Score 1437; DB 1; Length 1440;
 Best Local Similarity 43.5%; Pred. No. 1.9e-87;
 Matches 284; Conservative 133; Mismatches 198; Indels 38; Gaps 8;


```

Oy 74 IEFHGVEMLEVAFTFLSGSF-FENACLFVDGFLASRTFLVGHGLTIADIAMSNLGIG 132
Db 1 MEHEIDHMLEFSATKLSGSCDSTSTINELNCLSLRTLVGNLSLADLCVMATLTKGNA 60
Oy 133 QRMESLRKSKYQNLVRWFNSIDSEYKALNEVVAAPGKKGIGSPAPSLKEKHSD 192
Db 61 AMOQOLKOKKAPVHKWPFLEAQ-QAFQSV-----GTRKMDSTTKA 102
Oy 193 PSASE-----VDLPAGYGVCVAPAPSPGYLTHIGHAKAALLNKYAPERYOGRLLV 244
Db 103 RVAPEKQDVGKFWELPBAEMKQTVRRPPEASGTLHGHAKAALLNQHYNFNGKILIM 162
Oy 245 RPDNTNPKESNEFVENLLKDIETLGIKYDAVYTSDFPKLMEVASLTKQKAYIDT 304
Db 163 RPDNTNPKESNEFVENLLKDIETLGIKYDAVYTSDFPKLMEVASLTKQKAYIDT 222
Oy 305 PKQQRKRMGIESRCRNNTVEENLSLWKEMVNGTEGMOCCVARGKLDMDPNSKSLDP 364
Db 223 PAQWQKAREORIEBKRNKPIEKVLQWMEKKGSGHSCCLRAKIDMSNNGCMKDP 282
Oy 365 VYRCNTDPHHRVGSKYVYPTDACPVDALSGVTALRSSEYHDNAOYRRLDMG 424
Db 283 TLVCKIOPHRTGKKNVPTYPDACPVDALSGVTALRSSEYHDNAOYRRLDMG 342
Oy 425 LRARVEIYEPSRLNVTLLSKRKLMEVQNKKBEDMDPREPTVQGIYARGLYEALIQF 484
Db 343 IRKPYIWEYSRLNVTLLSKRKLMEVQNKKBEDMDPREPTVQGIYARGLYEALIQF 402
Oy 485 ILQOGASNLNLMEMDKLMTINKIIDVCAHHTAVLKQORVIFLTNGPEEPYRILIPR 544
Db 403 IAAQSSSRVNMEDKIMANRKYIDVAPRYVALLKKEVYI PVVPEAOES--MKEVAK 460
Oy 545 HKKEGAGKXATTPANRIMLDYADAANKGEVTLTMGNALIVEI-KVSSGVTELVG 603
Db 461 HPKAPVGLKAVWTSPPKVISGADSETSESGEMVTFINWGNLNTKIKHNAIDGKILSIDA 520
Oy 604 ELHLEGSVYTKTKITWLMADIEELVPLSL--VEPDYILSKKKLEDEDFLNLNPTCRR 661
Db 521 KFNLENDYKKTITVWLAETTHALPIVICYTHLITKPVLGDEDFEFOYVANKNSGHE 580
Oy 662 IPALGDANMRNIKGEIILQERKGYRCD-----APFIRSSKPVLPALPDG 708
Db 581 ELMIGDPLCKDLKGGDIQLORGFICDQYEPVSPYSCKEAPCVLLIYIPDG 633

RESULT 7
T27567
hypothetical protein ZC434.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27567
R:Wilkinson, J.
submitted to the EMBL Data Library, July 1996
A:Reference number: Z20388
A:Accession: T27567
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Reids: 1-1149 <WTL>
A:Cross-references: EMBL:Z75714; PIDN: CAB00060.1; GSPDB: GNM00019; CESP: ZC434.5
A:Experimental source: clone ZC434
C:Genetics:
A:Gene: CESP:ZC434.5
A:Map position: 1
A:Introns: 47/2; 502/2; 533/3; 649/3; 1110/3

Query Match 35.9%; Score 1339; DB 2; Length 1149;
Best Local Similarity 39.5%; Pred. No. 4.7e-81;
Matches 292; Conservative 132; Mismatches 243; Indels 72; Gaps 13;

Oy 2 EAALSPKSDSPISITICAAKLVGLPLTINSLAAGAPTLQFA-----SGESLHGV 52
Db 6 ELVIVKANREQPPVYSIILALASGFSL-----EKSVQFSEKQALNLMIDELLISND 55

```

```

Oy 53 NPILLYIARGASIA-SLSGKNDIEFGHVEWLEBAP-----TFLSGSEFENACLFVD 103
Db 56 VEIARIIAQSTADADSLGSSIIDPAVVDLTNLTAVTKNDYSLLGKDFP----- 108
Oy 104 GFLASRTFLVGHGLTIADIAMSNLGIGQRMESLRKSKYQNLVRWFNSIDSEYKALN 163
Db 109 -----TKILDNSTLVDAFALFS-----VAHNPOLK-----AKSGIIDLKYLKEPTL 150
Oy 164 EVNAAPYKRGIGKSP--APSLKEKHSDSPAPEDVLPAGYGVCVAPAPSPGYL 220
Db 151 AAANPFGLYSAAPATATAGTKK-----KDEGKFWELPBAEMKQTVRRPPEASGYL 206
Oy 221 HIGHAKAALLNKYAPERYOGRLLVIRPDNTNPKESNEFVENLLKDIETLGIKYDAVYTS 280
Db 207 HIGHAKAALLNQQYQAFEGQLIRFPDNTNPKESNAHEHYIKDLSMLNLTVPBRWTHSS 266
Oy 281 DYFPKLMEVASLTKQKAYIDTDPKQQRKRMGIESRCRNNTVEENLSLWKEMVNGT 340
Db 267 DHFEMLLTMCEKILKEKAFVDDTDTETMRNERORODSRNRSNTPEKNILQWMEKKG 326
Oy 341 ERGQCCVARGKLDMDPNSKSLRDPVYRCNTDPHHRVGSKYVYPTDACPVDALSGV 400
Db 327 PKGLTCCVRMKIDMKSNNGARDPITYRCPEEYVTKLKYVPTDFTCPIDVSEGV 386
Oy 401 THALRSSEYHDNRNAYRRIIDQMLRAREIYEFGRNLNVTLLSKRKLMEVQNKKBEDW 460
Db 387 THARTTEYHNRDQYIFICALGRPHIMEYARLNTNTVMSKRLTWVDGSHBGW 446
Oy 461 TDPEPFLVQIVRRGLKVEALIQFLQOGASKNLNMEMDKLMTINKIIDVCAHHTAV 520
Db 447 DDPLPLTVARGVRRGLTVEGKQFIVAQSGSRVNMEDKIMANRKYIDVAPRYVAL 506
Oy 521 LKQORVIFLTNGPEEPYRILIPRHKKEGAGKXATTPANRIMLDYADAANKGEVY 579
Db 507 DSTPLVSLTSDISDTSNV-SLHPKNAEISGKDVHKRKLLEQVDAALKEGELVT 565
Oy 580 LMDGNALIVEIKESGVTELVGELHLEGSVYTKTKITWLMADIEELVPLSLVEF 635
Db 566 FVMGNKIKIGIEKGAVITKISTLDLNDTYKTKTVTLGAVKABAGTIVVYADY 625
Oy 636 DYILSKKKLEDEDFLNLNPTCRRRIIPALGDANMRNIKGEIILQERKGYRCDAPFIR 695
Db 626 DHISKALIGDEBWKQFINDSVHYTKMGEPALIKVYKGGDIQLQIRKGYIYDQFPNP 685
Oy 696 SSK-----PVLPALPDG 708
Db 686 KSELGVTPLLLAIPDG 704

RESULT 8
AC0585
glutaminyl-tRNA synthetase [imported] - Salmonella enterica subsp. enterica serovar Typh
C:Species: Salmonella enterica subsp. enterica serovar Typh
A:Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AC0585
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,
th, T.; Comerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov.
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AC0585
A:Status: preliminary
A:Molecule type: DNA
A:Reids: 1-555 <PAR>
A:Cross-references: GB:AL513382; PIDN: CAD05149.1; PID: gl6501922; GSPDB: GNM0176
C:Genetics:
A:Superfamily: glutamine-tRNA ligase; glutamine-tRNA ligase homology

Query Match 20.3%; Score 757.5; DB 2; Length 555;

```

Best Local Similarity 35.1%; Pred. No. 1e-42;
Matches 189; Conservative 95; Mismatches 201; Indels 53; Gaps 15;

Qy 199 DLPAKVGKVCVRPAPEPSGYLHGHAKALINKYFAERYQGRLLVPRDDTPNPKESNEF 258
19 DLASGKHTTHTPRPPENGYLHGHAKSICLNFGIADYQGCNLRFDITNPVKEIDIEY 78
Db 259 VENLKDIETIGIKYDA-VTYSDFPKLMEASBLIKOGAKYIDDTPEQMRKER---- 313
79 VDSIKNDVEMLGFHWSGDIRYSSDFDQHAHYAVALINKGLAYDELTPEDIREYRGLT 138
Qy 314 MDGIESRCRNNTVEENLSLWKEM-VNGTERGMOCCVVRGKLDMDPNKSLRDPVYRCNTD 372
139 APGKNSPRDRSVEENLALFEKMRGFEBC-KACLRKAKIDMASPFYMRDPLVLRIFA 197
Qy 373 PNRVGSKYKVPYPTDPAFCPPVDALBGTVALRSSEYHNRNAOYRRLIDQMLR-RVEIY 431
198 EHHQGTGKWCYIPWDFHNCISDALBGTISLCTLEFPQDNRLYDWDVLDNTITIPHPROX 257
Qy 432 EFSRLNMYTLLSKRKLIMFVONKVEDWTPRPFTVOGIVRGKLYEALIQFILOQAS 491
258 EFSRLNMYTLLSKRKLIMLVTDKVIEGMDPRMPTISGLRRGYTAASIREFCRIGVT 317
Qy 492 KNLNLMEMDKLMTINKKIIDPVCARHTAVLKDQRIFFLTNGPRE-EPPVRLIPRHKKEG 550
318 KQDNTIEASLESICREBLNENAPRAMAVIDPVKLV-TEVPQGESEVMTNPNPKPE 375
Qy 551 AGKATTPANRIMLDYAD-AAALNK-----GEEVTLMQNA-IVKEIKVESGVTTEL 601
376 MGSREVPSPGSEIWDRAFRERANKQYRLVWGKEVRRL--NAVYIAERVEKDA----- 428
Qy 602 VGEHLGSGVKT-----TKLK--ITWLADEIEVLPSLVEFDYLISK 642
429 -----EGNITITICTYDADTLISKPADGRKVGVIHWVS--VAHALPIEIRLYDLFVSVP 481
Qy 643 KLEDEDFLDNLNPTCRREIPALGDANMRNIRKGGIIOLERKGYRCAPFRSSKPV 700
482 NPGAEDFLSVINPESLVIKOGFAPSLKAVAKAFQEREGYCLDSRYATADKLV 539
Db

RESULT 9
glutamine-tRNA ligase (EC 6.1.1.18) [validated] - Escherichia coli (strain K-12)
N/Aternate names: glutaminyl-tRNA synthetase
C/Species: Escherichia coli
C/Date: 13-Jun-1983 #sequence revision 05-Dec-1997 #text_change 03-Jun-2002
C/Accession: G64802; A92346; S03376; A01190; A31223
R/Bletter: F.R., Plunkett III, G., Bloch, C.A., Petina, N.T., Burland, V., Riley, M.; CC
.A.; Rose, D.U.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MUID:97426617; PMID:9278503
A/Accession: G64802
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-554 <BLAT>
A/Cross-references: GB:M10187; GB:U00096; NID:91786888; PIDN:AAC73774.1; PID:91786895;
R/Experimental source: strain K-12, substrain MG1655
R/Yamao, F.; Imokuchi, H.; Cheung, A.; Ozeki, H.; Söll, D.
J. Biol. Chem. 257, 11639-11643, 1982
A/Title: Escherichia coli glutaminyl-tRNA synthetase.
A/Reference number: A92346; MUID:83007336; PMID:6288695
A/Accession: A92346
A/Molecule type: DNA
A/Residues: 2-548, 'GR', 551 <YAM>
A/Cross-references: GB:M10187; NID:9146174; PIDN:AAA23884.1; PID:9146175
R/Uemura, H.; Conley, J.; Yamao, F.; Rogers, J.; Soell, D.
Protein Seq. Data Anal. 1, 479-485, 1988
A/Title: Escherichia coli glutaminyl-tRNA synthetase: a single amino acid replacement re-
A/Reference number: S03376; MUID:89113343; PMID:2464170
A/Accession: S03376
A/Molecule type: DNA
A/Residues: 2-554 <UEM>

R/Hoben, P.; Royal, N.; Cheung, A.; Yamao, F.; Blemann, K.; Söll, D.
J. Biol. Chem. 257, 11644-11650, 1982
A/Title: Escherichia coli glutaminyl-tRNA synthetase. II. Characterization of the glnS g
A/Reference number: A92347; MUID:83007237; PMID:6798844
A/Contents: annotation; confirmation of amino and carboxyl ends by amino acid analysis;
C/Genetics:
A/Gene: glnS
A/Map position: 15 min
A/Function:
C/Description: EC 6.1.1.18 [validated, MUID:89113343]
A/Pathway: protein biosynthesis
C/Superfamily: glutamine-tRNA ligase; glutamine-tRNA ligase homology
C/Keywords: aminocyl-tRNA synthetase; ATP; ligase; protein biosynthesis
F/28-310/Domain: glutamine-tRNA ligase homology <BDL>

Query Match 20.2%; Score 753; DB 1; Length 554;
Best Local Similarity 35.2%; Pred. No. 2e-42;
Matches 189; Conservative 91; Mismatches 205; Indels 52; Gaps 14;

Qy 199 DLPAKVGKVCVRPAPEPSGYLHGHAKALINKYFAERYQGRLLVPRDDTPNPKESNEF 258
19 DLASGKHTTHTPRPPENGYLHGHAKSICLNFGIADYQGCNLRFDITNPVKEIDIEY 78
Db 259 VENLKDIETIGIKYDA-VTYSDFPKLMEASBLIKOGAKYIDDTPEQMRKER---- 313
79 VDSIKNDVEMLGFHWSGDIRYSSDFDQHAHYAVALINKGLAYDELTPEDIREYRGLT 138
Qy 314 MDGIESRCRNNTVEENLSLWKEM-VNGTERGMOCCVVRGKLDMDPNKSLRDPVYRCNTD 372
139 APGKNSPRDRSVEENLALFEKMRGFEBC-KACLRKAKIDMASPFYMRDPLVLRIFA 197
Qy 373 PNRVGSKYKVPYPTDPAFCPPVDALBGTVALRSSEYHNRNAOYRRLIDQMLR-RVEIY 431
198 EHHQGTGKWCYIPWDFHNCISDALBGTISLCTLEFPQDNRLYDWDVLDNTITIPHPROX 257
Qy 432 EFSRLNMYTLLSKRKLIMFVONKVEDWTPRPFTVOGIVRGKLYEALIQFILOQAS 491
258 EFSRLNMYTLLSKRKLIMLVTDKVIEGMDPRMPTISGLRRGYTAASIREFCRIGVT 317
Qy 492 KNLNLMEMDKLMTINKKIIDPVCARHTAVLKDQRIFFLTNGPRE-EPPVRLIPRHKKEG 551
318 KQDNTIEASLESICREBLNENAPRAMAVIDPVKLV-TEVPQGESEVMTNPNPKPE 375
Qy 551 AGKATTPANRIMLDYAD-AAALNK-----GEEVTLMQNA-IVKEIKVESGVTTEL 602
376 MGSREVPSPGSEIWDRAFRERANKQYRLVWGKEVRRL--NAVYIAERVEKDA----- 427
Qy 602 VGEHLGSGVKT-----TKLK--ITWLADEIEVLPSLVEFDYLISK 643
428 -----EGNITITICTYDADTLISKPADGRKVGVIHWVSAAHAL-PEIRLYDLFVSVP 481
Qy 644 LEEDEDFLDNLNPTCRREIPALGDANMRNIRKGGIIOLERKGYRCAPFRSSKPV 700
482 NPGAEDFLSVINPESLVIKOGFAPSLKAVAKAFQEREGYCLDSRYATADKLV 538
Db

RESULT 10
glutamine tRNA synthetase [imported] - Escherichia coli (strain O157:H7, substrain R1MD
F90717
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C/Accession: F90717
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hatboro, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno-
A/Reference number: A99629; MUID:21156231; PMID:11258796
A/Accession: F90717
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-554 <HAY>
A/Cross-references: GB:BA000007; PIDN:BA834133.1; PID:913360168; GSPDB:GN00154
A/Experimental source: strain O157:H7, substrain R1MD 0509952

C;Genetics:

A;Gene: ECS0710

C; Superfamily: glutamine-tRNA ligase; glutamine-tRNA ligase homology

Query Match	20.08; Score 747; DB 2; Length 554;
-------------	-------------------------------------

Best Local Similarity 34.8%; Pred. No. 5.1e-42;

Matches 187; Conservative 93; Mismatches 205; Indels 52; Gaps 14;

QY 199 LPLGAKYGVKCVRFAPESPGLYHTIGHAQAALLNKKYFAARYGGRLLVRFDDTNPKESNEF 258
 Db 19 DLASGKHTVTRPPPEPBGYLHTIGHAKSICLNFGIADYGGQCNLRDDTNPKEDLEY 78
 QY 259 VENTLKDIETIGIKYDA--VITYSDYFPFKLMEASBLIKOGKAYIIDTPEQMKRER---- 313
 Db 79 VDSIKNDIEWIGFHMSGNVRYSSDYFDQLHAYALIELINKGLAYDELPEDIRYRGILT 138
 QY 314 MDGIESRCRRNTVBEENISLTKEM--VNGTERGMQCCVRGKLDMDQPNKSLRDPVYRCNTD 372
 Db 139 QPGKNSPRDRSVEENLALPEKMRGTGFEEG--KACLRAKIMASPFIVMRDPVLYRIKFA 197
 QY 373 PHHVGSKKYKYPYDFAPCEFDVDALEGVTHARSEYHEDRAQYRILODGLR-RVEIY 431
 Db 198 EHHOTGNKWCIPYMDVDFHCISDLBEGITHSLCTLEFODNNRLLYDWLNDNTTIVHPROY 257
 QY 432 EFSRLNMYVYTLISKRKLIMFYQNNKKEVDWPTPREFTYVGIVRGRGLKYEALIQTLIOGAS 491
 Db 258 EFSRLNLEYTWMSRKRLNLVTDGKHEGMDPRMPTISGLRRRGYTAASIEFCRIGVT 317
 QY 492 KNLMLMMDKLMWTINKKIIDPVCARHTVALKDQRYIFTLTNGPEEPFVRIIPRHKKEGA 551
 Db 318 KQDWTIEMASLESCEIRDLENAPRAMAVIDPVKLTIVENYQEGE--MVTMHPNPKEM 375
 QY 552 GKRAATTFNRIWLYAD--AAAINK-----GEEVTLMDQNA--IYVEIKYESGIVITELV 602
 Db 376 GSRQVPSGEIWIIDRSDBREANQYKRLVZGKERLE--NAVYIAERVEKDA---- 427
 QY 603 GELHLEGSVKT-----TKLK--ITWLADIEELVPLSLVEFDLISKK 643
 Db 428 -----EGNITTIFCTYDADLTLSKOPADGRKTKGVYHWSAHL--PVEIRIYDLTFSVPN 481
 QY 644 LEEDBDPLDINPCTRREIPALGDANMNRKGEITIQLEKGYRCDAPIRISKPV 700
 Db 482 PGAADDFSVINPESLVIKOGFAEPESLKDVAAGAKAFOPERGEGYCLDSRHSIAKPV 538

RESULT 11
 F85567
 glutamine tRNA synthetase [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: F85567
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobbeck, E.J.; Davis, N.W.; Llim, A.; Diallantha, E.; Potamoultis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: F85567
 A:Status: preliminary
 A:Molecule type: DNA
 A:Restrictions: 1-554 <STO>
 A:Cross-references: GB:AE005174; NID:G12513581; PIDN:AAG55002.1; GSPDB:GN00145; UMG:P.2008
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: glsB
 C:Superfamily: glutamine-tRNA ligase; glutamine-tRNA ligase homology

Query Match 20.0%; Score 747; DB 2; Length 554;
 Best Local Similarity 34.8%; Pred. No. 5.1e-42;
 Matches 187; Conservative 93; Mismatches 205; Indels 52; Gaps 14;

199 DLPGAKYGVKCVRFAPESPGLYHTIGHAQAALLNKKYFAERYGGRLLVRFDDTNPKESNEF 258
 19 DLASGKHTVTRPPPEPBGYLHTIGHAKSICLNFGIADYGGQCNLRDDTNPKEDLEY 78

```

QY 259 VBNLKIDLETIGIKYDA-VYTTSDYPRFKLMAASLJKOGXAYIDDTPEKQMKRER---- 313
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 79 VDSIKNDVEMJGFHMSGVNRYSDDYFDQLAHYAELELNKGLAYADELTPEQIREGRGTLT 138
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 314 MDGIESRCRNNTVEENLSLWKEM-VNGTERGMQCCVBRGKLDMOOPNKSJRPVYYRCNTD 372
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 139 QPGKNSPYRDSVSEENLALFEKMRTOGSEEB-KACSLAKIDMAEPFIYMRPVLIRLIFA 197
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 373 PHHRVGSRYKYQPYDYDFACPFVDALBEGVTALRSSEYHNRNNAQYRRILODMGLR-RVEIY 431
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 198 EHHQGTGKMKCIYPMWDFTHCISDALBEGITHSLCTLEBFQDNRLVDWVLDNITIEVHPROY 257
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 432 EFSRLNMYVTLTSKRKLMLFVQNKKVEDMDTPREFPTQVIGYRGLKVEALIOFLIOQOAS 491
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 258 EFSRLNLEYTMSGRKLNLLVTDKHNBEGMDPRPFTISGLRRRGYTAASINIEFCKRISQVT 317
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 492 KNLNLMEMWDLMTINKKIIDPVCARHTVALYKQGVITFTLNGPEEPFVRIILPRHKKEGA 551
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 318 KQDWTITIEASLESICREDPLENENAPRAMAVIDPVKLVIENYQSEGE--MVTNPHNPKBEM 375
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 552 GKXATTFANRWLWLYAD-AAALNK-----GESEVTLMDMGN-AIYEIKYESGVITELY 602
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 376 GSRQVPSFGEIWIHDRSDREBEANKQRYKLVLGKEVRLR--NAVYITAEERKEKA----- 427
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 603 GELHLEGSVKT-----TKLK--ITWLADIIEELVPLSLVEFDYLSIKKK 643
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 428 -----EGNITITICTYDADLTLSKDPADQKRYKGVIIHWVSAHAL-PVEIRLYDLRFSVBN 481
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 644 LEEBBDPFLDNINPCTRREIIPALGDAMNRNIRKGETIQERKGYRCDAPIRFSKQPV 700
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 482 PGAADDPFLSVINPSSLVIYKQGFAPPSLKDAVAGAFQFEREGGYCCLDSRHSHTAKPV 538
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

RESULT 12
 AB0321
 glutamine-tRNA ligase (EC 6.1.1.18) [imported] - Yersinia pestis (strain CO92)
 C:Species: Yersinia pestis
 C:Date: 02-Nov-2001
 C:Accession: AB0321
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
 geno-Tarrag, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
 Nature 413, 523-527, 2001
 A>Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AB0321
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-555 <KIR>
 A:Cross-references: GB:ALJ590842; PIDN:CAC92873.1; PID:g15980617; GSPDB:GN00175
 C:Genetics:
 A:Gene: glus
 C:Superfamily: glutamine-tRNA ligase; glutamine-tRNA ligase homology
 C:Keywords: ligase

Query Match	19.9%	Score 743.5	DB 2	Length 555
Best local similarity	35.8%	Pred. No. 8.7e-42		
Matches 196	Conservative 80	Mismatches 214	Indels 43	Gaps 15
199	DLPGAKYGVCVRPAPESPGLYHIGHAKAALLNFKYFAERYGRLIVFDDTNPSEKSENF	258		
Db	19 DLAGKHTSVHTRPPEPENGYLIGHAKSICLNGIAEDYGGCCNLRDPTNPVKEDEVF	78		
QY	259 VENLLKDIETLTGITYDA-VYTTSDYFPKLKEMASLTKQKAYIDDTLPKEQMKER----	313		
QY	79. VESIKRDVEMIGFVMSGDVRYSSDPOLVOYAVELINKGLAYVDELTPKEQMKREYRGTLT	138		
QY	314 MDGIESRCNNVTVENSLWKEMVNWGTERGQCVCVRGLDKWDQNKSLRPDVYRCNTDP	373		
Db	139 APGNKSPYRDSVEENLALFEKMRAGPABSTACLRPAKIDPASPFIWRDDVLYIRKFAE	198		
QY	374 HHRGSKYKVPYTDACFPVDALGVTHAIRSSEHYDRNAYRILODMGLR-RVEIYE	432		

Db 199 HHGSGNKKICITPMWDFTHCTICIDALEGITHSLCTLEFQNNRRLYDVAVLNDNISIDCHPRXYE 258
Qy 433 FSRILNMTYLLSKRKLTFVONKXVEDWTPDPPTVQIVRGKVEALIOFILOOGASK 492
Db 259 FSRILNMTYLLSKRKLTFVONKXVEDWTPDPPTVQIVRGKVEALIOFILOOGASK 318
Qy 493 NLNLMEMKMTINKKIIDPCARHTAVLKQORVIFLTNLNGPEEPFVALILPRHKKFEBAG 552
Db 319 QNNVEMMSLSCIRDDINENHAPRAMAVLDPKIVTENRAAGEEMLT--MPVHPNPMG 376
Qy 553 KKATTPANRIMLDYAD-AAALNK-----GEEVTLMDGNA-LYKEIKVE---SGVITE 600
Db 377 SRQVPFSEIYIIRADPREBANKQYKRLVLGKEVRLR---NAVYKAERVKDANGNTT 433
Qy 601 LVGELHLE-----GSVTKTKLKITWLDIEBLVPLSVEFDYLSKKLLEDEDFLNL 654
Db 434 LYCSYDAETLNKDPADGRKVKGVIMWVS--VAHALPAEIRLYRDLFPNVNPPAAEDFLSTI 492
Qy 655 NP---CTPREI--PALGDANMRNIRGELIQERKGYRCDAPIRSKPVYL 702
Db 493 NPESLVIHQGFVEPSLADA-----VSDKTYQFEREGYFCADS--RYSRPGAL 537

RESULT 13

H82254
glutaminyl-tRNA synthetase VC0997 [imported] - Vibrio cholerae (strain N16961 serogroup
C/Species: Vibrio cholerae
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C/Accession: H82254
R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Diegel, I.; Sellere, F.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A/Reference number: A62035; MUID:2040683; PMID:10952301
A/Accession: H82254
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-556 <HEI>
A/Cross-references: GB:AE004181; GB:AE003852; NID:99655454; PIDD:AA94158.1; GSPDB:GN001
A/Experimental source: serogroup O1, strain N16961, biotype El Tor
C/Genetics:
A/Gene: VC0997
A/Map position: 1
C/Superfamily: glutamine-tRNA ligase; glutamine-tRNA ligase homology

Query Match 19.3% Score 721; DB 2; Length 556;
Best Local Similarity 34.3%; Pred. 2.8e-40;
Matches 185; Conservative 91; Mismatches 208; Indels 56; Gaps 13;

Qy 199 DLPGAQKVCVRAPEPSGYLHGHAKAALLNKYFAERYGRLIVRPDDTNPSKESNEF 258
Db 19 DLADGKHTVTRFPEPENGYLHGHAKSICLNFGIADYQOQCNLRFDDTNPEKENDLEY 78
Qy 259 VENLKDIETIGIKYDA-VTYSDFPKLMEASBLIKOGKAYIDDPTEQMRKER----- 313
Db 79 VESIKKDVTLWGFMDGSEGCYSSDYFDKLYEYALIELIKGLAYVDELPEQIRREYRGLT 138
Qy 314 MDGIESRCNNVTENLSLMEKMNVTGERMOCQVRGLDMQDPNKSIRDPVYRCNTDP 373
Db 139 EPGKISPYRDRSVEENLALPEKMRAGEFAEGACIRAKIDMASSTVWMDPVLVVRPAE 198
Qy 374 HHRVGSKKVPTYPDFACPFVDALEGVTHALRSSEYHDMNAQYRIILDMGLR-RVEIYE 432
Db 199 HHQGTGDKKICIPMYVFTHCISDALEGIHSTICLFQNNRRLYDVAVLNDNITTPCHPRXYE 258
Qy 433 FSRILNMTYLLSKRKLTFVONKXVEDWTPDPPTVQIVRGKVEALIOFILOOGASK 492
Db 259 FSRILNMTYLLSKRKLTFVONKXVEDWTPDPPTVQIVRGKVEALIOFILOOGASK 318
Qy 493 NLNLMEMKMTINKKIIDPCARHTAVLKQORVIFLTNLNGPEEPFVALILPRHKKFEBAG 552

Db 319 QENNIEYSALES CIRDDLENENAPRAMAVLDPVKLVIEENFAGVETITLANHPNKKPEM-- 376
Qy 542 LPRHKKFEGAKKATTPANRIMLDYAD-AAALNK-----GEEVTLMDGNA-LYKEIKV 593
Db 377 -----GDRVPTRELMIREDPREBANKQYKRLVLGKEVRLR--GAVYKAERI 424
Qy 594 ---ESGVIT-----ELVGEHLGSGSVTKTKLKITWLDIEBLVPLSVEFDYLSKK 642
Db 425 EKDEQGNITTFGCVDEPDLCKNPADG--RKVKGVIMWVS-AEKGVAPREFRYRLFTVP 481
Qy 643 KLEDEDFLNLNCTPREIIPALGDANMRNIRGELIQERKGYRCDAPIRSKPVYL 702
Db 482 NPGADNFAETINDESIVKQGVPEPSLVEAKPEFGYQFERMGYFCADN---KDSFGAL 538

RESULT 14

G64118
glutamine-tRNA ligase (EC 6.1.1.18) - Haemophilus influenzae (strain Rd KW20)
N/Alternate names: glutaminyl-tRNA synthetase
C/Species: Haemophilus influenzae
C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 03-Jun-2002
C/Accession: G64118
R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.
A/Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A/Reference number: A64000; MUID:95350630; PMID:7542800
A/Accession: G64118
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-557 <TIGR>
A/Cross-references: GB:U32814; GB:L42023; NID:91574809; PIDD:AA23001.1; PTD:91574816; T
A/Genetics:
C/Superfamily: glutamine-tRNA ligase; glutamine-tRNA ligase homology
C/Keywords: aminocacyl-tRNA synthetase; ligase; protein biosynthesis
F:36-319/Domain: glutamine-tRNA ligase homology <EBI>

Query Match 19.0% Score 710.5; DB 2; Length 557;
Best Local Similarity 34.5%; Pred. No. 1.4e-39;
Matches 182; Conservative 93; Mismatches 214; Indels 39; Gaps 13;

Qy 199 DLPGAQKVCVRAPEPSGYLHGHAKAALLNKYFAERYGRLIVRPDDTNPSKESNEF 258
Db 27 DLASGKHSVTRFPEPENGYLHGHAKSICLNFGIADYQOQCNLRFDDTNPEKENDLEY 86
Qy 259 VENLKDIETIGIKYDA-VTYSDFPKLMEASBLIKOGKAYIDDPTEQMRKER----- 313
Db 87 VDSIKADVWLGFKMGEPRYASDYFDALYGYAVELIKGLAYVDELSPEDEKREYRGLT 146
Qy 314 MDGIESRCNNVTENLSLMEKMNVTGERMOCQVRGLDMQDPNKSIRDPVYRCNTDP 373
Db 147 EPGKISPYRDRSVEENLALPEKMRAGEFAEGACIRAKIDMASSTVWMDPVLVVRPAE 206
Qy 374 HHRVGSKKVPTYPDFACPFVDALEGVTHALRSSEYHDMNAQYRIILDMGLR-VIY 431
Db 207 HHQGTGDKKICIPMYVFTHCISDALEGIHSTICLFQNNRRLYDVAVLNDNITTPCHPRXYE 258
Qy 432 EFSRLNMTYLLSKRKLTFVONKXVEDWTPDPPTVQIVRGKVEALIOFILOOGASK 491
Db 267 EFSRLNMTYLLSKRKLTFVONKXVEDWTPDPPTVQIVRGKVEALIOFILOOGASK 326
Qy 492 KNLNLMEMKMTINKKIIDPCARHTAVLKQORVIFLTNLNGPEEPFVALILPRHKKFEGA 551
Db 327 KQDNVVEYSALACRDLNENAPRAMAVLDPVKLVIEENFAGVETITLANHPNKKPEM-- 383
Qy 552 GKATTPANRIMLDYAD-AAALNK-----GEEVTLMDGNA-LYKEIKVE---SGVIT 599
Db 384 GERQLPFKEKLYIIRADPREBANKQYKRLVLGKEVRLR---NAVYKAERVKDANGNTT 440
Qy 600 -----ELVGEHLGSGSVTKTKLKITWLDIEBLVPLSVEFDYLSKKLLEDEDFL 651

Db 441 TIFCTYDPEITGKNPAG--RKVGVIHWSAVNN-HPAEFRILYRLFTVNPGEADDIE 497
QY 652 DNLPCTRREIIPALGDAMRNIRKGEIIQLRKGYRCDAPIRSSKP 699
Db 498 SVLNPNSLIVIKQGFVEQSLANAERKGYQFEREGYFCADS--KDSRP 542

Search completed: January 25, 2004, 17:14:35
Job time : 72 secs

RESULT 15

F83421

glutaminyl-tRNA synthetase PA1794 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: F83421

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,

J.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: F83421

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-556 <STO>

A:Cross-references: GB:AE004605; GB:AE004091; NID:99947771; PIDN:AA05183.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: glsB; PA1794

C:Superfamily: glutamine-tRNA ligase; glutamine-tRNA ligase homology

Query Match 18.9%; Score 704.5; DB 2; Length 556;
Best Local Similarity 35.3%; Pred. No. 3.5e-39;

Matches 191; Conservative 81; Mismatches 216; Indels 53; Gaps 17;

QY 197 EVDLPAGKVGKVCVPAPDPSPGYLHGHAKALINKYFAERYQGRIVRPDDTNPSESN 256
Db 18 QADIDAGKHAIVTRFPPEPNGLYHGHAKSICLNFGLAQFAGDCHLFPDTPAKEDQ 77
QY 257 EFVENLKDITELGKYNA-VTYSDFPKLMEAMESLTKOGKAVIDTTPKEQKER-- 313
Db 78 EYIDALEADIDWLGQWGSVCYASNYFDQLHAWAVELIKAGKAFVCDLGPPEMERXYRGT 137
QY 314 --MDGISRCRNNITVEENLSLMEKVNNGTERGMQCCVGLMDOPNKLDPVYYRCNT 371
Db 138 LTERGRNSPYRDRSVEENLDLPARKAGFPDGRASLPAKIDMGSPNNMLDPITYRIH 197
QY 372 DPHRVGSKYKVPYTPDFACPFVDALGCVTHALRSSEYHNRNAQYRILLQDGL-RRVEI 430
Db 431 YEFRLMNVYTLISKRLIMVQNKVVDWTDPRPFTVQGIVRGLKVEALIQPLQGA 490
QY 198 AHHTQGDKWCICYPSYDPTHGSDALEGITHSICLTFEDHRPLYEMFLMLPVPAQPRQ 257
Db 258 YEFRLMNVYTLISKRLIMVQNKVVDWTDPRPFTVQGIVRGLKVEALIQPLQGA 490
QY 491 SKNLNLEWMDLMTINKKIIDPVCARHTAVLKQDQVIFLTNGPEPEFVRI-LPRHKKFE 549
Db 318 NRAGCVDIGWLEFSIRHDLATAPRAMCVLPLKV--ITNYPEGQVENLELPRHK-E 374
QY 550 GAGKATTFANRIMLDVADAAN-----KGEVTLMDGNAIVK--EIKVESGVI 598
Db 375 DMGVRVLPFGRELFDADGFEVVPAGYKRLIPGGEVRLR--GSYVIRADEALIKADGNI 432
QY 599 TEL-----VGEILLEGSVKTKITWLADIEELVPLSLVEFDYLI---SKKLEED 647
Db 433 VELRCSYDPDTLTK-NPEG--RKVGVIHWSV-PAEGSVCEVRVLDRLFRSANEKAEKG 488
QY 648 EDFLDNLP-----CTRREIPALGDAMRNIRKGEIIQLRKGYRCDAPIRSSKP 699
Db 489 GSFIDNINADSLQVLAGC--RAEBSLGQANPE-----DRQFEREGYFVADLKDSRPGKP 541
QY 700 V 700
Db 542 V 542

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2004, 16:58:16 ; Search time 39 seconds
(without alignments)

862.156 Million cell updates/sec

Title: US-09-831-683B-10

Perfect score: 3734
Sequence: 1 MEALSFKSDSPPIICAA.....SSKEVLFALPDGRQASLS 715

Scoring table: BIOSUM62.
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1574.5	42.2	716 1	SYEC_SCHPO
2	1562	41.8	724 1	SYEC_YEAST
3	1498	40.1	1714 1	SYEP_DROME
4	1437	38.5	1440 1	SYEP_HUMAN
5	757.5	20.3	554 1	SYO_SALT1
6	755.5	20.2	554 1	SYO_SALT2
7	753	20.2	553 1	SYO_SALT3
8	750	20.1	553 1	SYO_SALT4
9	747	20.0	553 1	SYO_SALT5
10	743.5	19.9	555 1	SYO_SALT6
11	734.5	19.7	556 1	SYO_SALT7
12	725	19.4	559 1	SYO_SALT8
13	721	19.3	556 1	SYO_SALT9
14	710.5	19.0	557 1	SYO_SALT10
15	705.5	18.9	552 1	SYO_SALT11
16	704.5	18.9	556 1	SYO_SALT12
17	690	18.5	571 1	SYO_SALT13
18	683.5	18.3	580 1	SYO_SALT14
19	682	18.3	580 1	SYO_SALT15
20	678	18.2	579 1	SYO_SALT16
21	677.5	18.1	579 1	SYO_SALT17
22	666.5	17.8	562 1	SYO_SALT18
23	666.5	17.8	579 1	SYO_SALT19
24	664.5	17.8	579 1	SYO_SALT20
25	653	17.5	571 1	SYE_PYRAB
26	651.5	17.4	572 1	SYE_PYRAB
27	636.5	17.0	794 1	SYO_LUPLU
28	624	16.7	570 1	SYE_PYRHO
29	594	15.9	809 1	SYO_YEAST
30	589.5	15.8	553 1	SYE_METTA
31	575	15.4	852 1	SYO_DEIRA
32	573	15.3	786 1	SYO_CABEL
33	572	15.3	811 1	SYO_SCHPO

34	561.5	15.0	775 1	SYO_HUMAN	P47897 homo sapien
35	559	15.0	551 1	SYE_ARCFU	O29979 archaeoglob
36	553	14.8	778 1	SYO_DROME	O9Y105 drosophila
37	548	14.7	566 1	SYE_SULTO	O97160 sulfolobus
38	540.5	14.5	571 1	SYE_METAC	O8T52 methanobact
39	537.5	14.4	553 1	SYE_METTH	O26157 methanobact
40	530	14.2	552 1	SYE_METTM	O50543 methanobact
41	529.5	14.2	571 1	SYE_METRA	O8Txb7 methanopyru
42	510.5	13.7	571 1	SYE_METMA	O8Pw52 methanobact
43	504	13.5	575 1	SYE_SULTO	P95968 sulfolobus
44	502.5	13.5	570 1	SYE_PYRAB	O8Zu33 pyrobaculum
45	480.5	12.9	548 1	SYE_THRAC	O9Yjms thermoplasm

ALIGNMENTS

RESULT 1	SYEC_SCHPO	STANDARD;	PRT;	716 AA.
ID	O13775;			
AC	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DE	Probable glutamyl-tRNA synthetase, cytoplasmic (EC 6.1.1.17)			
DE	(Glutamate--tRNA ligase) (GluRS).			
GN	SPAC17A5.15C.			
OS	Schizosaccharomyces pombe (Fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OX	NCBI_TaxID=4896;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=972;			
RX	MEDLINE=21848401; PubMed=11859360;			
RA	Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,			
RA	Sgouros K., Peat N., Hayles J., Baker S., Basham D., Bowman S.,			
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,			
RA	Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,			
RA	Gentles S., Gobie A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,			
RA	Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,			
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,			
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odel C.,			
RA	Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,			
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,			
RA	Skellern J., Simmonds M., Squares R., Stevens S., Stevens K.,			
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,			
RA	Woodward J., Voicakert G., Aert R., Robben J., Grymptre B.,			
RA	Welfjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,			
RA	Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,			
RA	Borzum K., Langer I., Beck A., Leirich H., Reinhardt R., Pohl T.M.,			
RA	Bozzyk P., Zimmermann W., Wedler S., Gloux S., Lelaure V., Mottier S.,			
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,			
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,			
RA	Lucas M., Rocher M., Galliard C., Tallada V.A., Garzon A., Thode G.,			
RA	Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,			
RA	Domiguez A., Revuelta J.U., Moreno S., Armstrong J., Forsburg S.L.,			
RA	Cerintu L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,			
RT	"The genome sequence of Schizosaccharomyces pombe."			
RL	Nature 415:871-880(2002).			
CC	-I- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +			
CC	diphosphate + L-glutamyl-tRNA(Glu).			
CC	-I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).			
CC	-I- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation-			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.ebi.ac.uk/announcements)			

or send an email to license@ib-sib.ch.

CC -----

CC EMBL, Z98849, CAB1515.1; -.

CC PIR, T37830, T37830.

CC HSSP, P00962, 1GTR.

CC GenDB, SPombe, SPAC17A5.15c; -.

CC InterPro, IPR004526, Gltx arch.

CC InterPro, IPR000924, Glu_tRNA-synt_1c.

CC InterPro, IPR001412, tRNA-synt_1.

CC Pfam, PF00749, tRNA-synt_1c; 1.

CC Pfam, PF03950, tRNA-synt_1c; 1.

CC PRINTS, PRO0987, TRNASYNTGLU.

CC TIGRfam, TIGR00463, gltx arch. 1.

CC PROSITE, PS00178, AA_tRNA_LIGASE_1; 1.

CC Aminoacyl-tRNA synthetase; protein biosynthesis; ligase; ATP-binding.

CC SITE 215 224 "HIGH" REGION.

CC SITE 441 445 "KMSKS" REGION.

CC BINDING 444 444 ATP (BY SIMILARITY).

CC SEQUENCE 716 AA; 80749 MW; 2ACE0A35ED393227 CRC64;

Query Match 42.2%; Score 1574.5; DB 1; Length 716;
 Best Local Similarity 49.5%; Pred. No. 1.1e-101;
 Matches 319; Conservative 102; Mismatches 201; Indels 23; Gaps 10;

QY 70 GNIDEPGHVWBLEYAPFLSGS-----EFENACLFVNGFLASRTFLVGHGTTIDIAVW 125
 DB 81 GENDBSL--VESWETA-SALAGNNHFLSLAQDLHLMRSLFVGYSLTSADFSIW 137
 QY 126 SNLAGIGQWESLSKSKYQNLVRFNSIDSEYKALNEVVAFAFGKGIKSPAPSLKE 185
 DB 138 GALKSNMNAAGAVR-TGQYFNLAIRYKIMDSQ--NAVETMEEFKAVANISK-----KQ 188
 QY 186 KVHDSKDSAPDEVLDPAKVGKVCVPAPEPSGYLHIGAKAALINKYFAERYQRLIVR 245
 DB 189 K---SSGPNY-ELGPDALIDGKVTVPPEPSGYLHIGAKAALINQFANKYHGLIYR 244
 QY 246 FDDTNPSSNEFEVNLKDIETLGIKYDATTYTSDDYFPKLEMAESLIGKQKATIDTP 305
 DB 245 FDDTNPSSKENSEFOALIEDVALLGIKRPDVYTSDDYDTLHQVCVDMIKSGQAAVADTD 304
 QY 306 KEOMKEMMDGIESRCNNTVENSLWKEWNGTERGMOCCVCRKLMODPNKSLRPV 365
 DB 305 VETMHERTEGIPSGHRDRPESLEIISMDKSGDVGLKNCIRAKIYEENPKAMRPV 364
 QY 366 YYRCNTDPMHRYGSKYKYPYDPAFCFVDALLEGVTHALRSSEYHADRNAOYRRILQDKL 425
 DB 365 IYRCNLLPHHRTGTYRAVPTDFACPIYDSLEGVTHALRTTEYDRNPLVQMMIKAMNL 424
 QY 426 RRVLEYSRLMNYTLLSKRCLMFVQKKYEDTDRFPVQGIYVRGLKVEALLQFI 485
 DB 425 RKIHWEBSRNMFPVTLISKRLTEIVDHGLWGMDDPRFPVRRGRRGMTIELQOYI 484
 QY 486 LQOGASKULNMEWKLWTINKKIIDPYCARHTALVDOORVFTLTNGPEEPFVILPRH 545
 DB 485 VSQGSKUILITDMTSPFATNKIIDPYAPRHTAVESGDVYKATLVNCPADYADRRRH 544
 QY 546 KKFEGAGKATTFANRIMLDVADAAANKGEVTLMDGNALVKEIYK-SGVITELVGE 604
 DB 545 KKNPELGKKSI FANEILLIEQADAQSPQDEEVLTLMDGNAAVRIINNDASGKVTSLKE 604
 QY 605 LHLBSGVTTKLKITWLADIELVPLSLVEFYLISKKLLEDEDFLNDLPCTREIPA 664
 DB 605 LHLDDPFKTEKVTWLADTEDEKTPVDLVDFYLLITKCLLEGENVYKDFLTPTQTEFHSV 664
 QY 665 LGDAMRNINKEGIIQLERKGYRCDAPFIRSKSVVLFAIPDGR 709
 DB 665 FADVGIRKNLKKGDIIQVERKGIYDVFP--DGTQAVLFNIDGK 707

RESULT 2
 SYEC YEAST
 ID SYEC YEAST STANDARD; PRT; 724 AA.
 AC P46655;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Glutamyl-tRNA synthetase, cytoplasmic (EC 6.1.1.17) (Glutamate--tRNA
 ligase) (GluRS) (P85).

GN YGL245W OR G0583 OR HRB724.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.

OX NCI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RA Frantz J D., Gilbert W.;

RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.

RC STRAIN=S288c;

RA Vanderbol M., Durand P., Portetelle D., Hilger F.;

RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 1-146 FROM N.A.

RC STRAIN=S288c / FY1679;

RA Colesac E., Maillier E., Neter P.;

RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
 diphosphate + L-glutamyl-tRNA(Glu).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.ibsib.ch/announce/>
 or send an email to license@ib-sib.ch).

CC -----

CC EMBL, U32265, AAA78805.1; -.

CC EMBL, Z49149, CAAB9009.1; -.

CC EMBL, Z72767, CAAB9694.1; -.

CC EMBL, X94357, CAAG4142.1; -.

CC PIR, S53934; S53934.

CC HSSP, P00962, 1GTR.

CC SGD, S0003214, YGL245W.

CC InterPro, IPR004526, Gltx arch.

CC InterPro, IPR000924, Glu_tRNA-synt_1c.

CC InterPro, IPR001412, tRNA-synt_1.

CC Pfam, PF00749, tRNA-synt_1c; 1.

CC Pfam, PF03950, tRNA-synt_1c; 1.

CC PRINTS, PRO0987, TRNASYNTGLU.

CC TIGRfam, TIGR00463, gltx arch. 1.

CC PROSITE, PS00178, AA_tRNA_LIGASE_1; 1.

CC Aminoacyl-tRNA synthetase; protein biosynthesis; ligase; ATP-binding.

CC SITE 226 235 "HIGH" REGION.

CC SITE 453 457 "KMSKS" REGION.

CC BINDING 456 456 ATP (BY SIMILARITY).

CC CONFLICT 225 225 E -> D (IN REF. 1).

CC CONFLICT 489 489 V -> A (IN REF. 1).

CC CONFLICT 526 526 P -> S (IN REF. 1).

CC CONFLICT 562 562 V -> M (IN REF. 1).

CC CONFLICT 714 724 GKSYNKGAKK -> VNLSTSMVQRNKHISNVTYLCYFS
 TSTP (IN REF. 1).

CC SEQUENCE 724 AA; 82662 MW; 34669BFB9CD41BE CRC64;

Query Match 41.8%; Score 1562; DB 1; Length 724;
 Best Local Similarity 45.7%; Pred. No. 8.1e-101;
 Matches 328; Conservative 121; Mismatches 212; Indels 56; Gaps 14;

QY 27 LTINHSIAAGAPTLPAGSGSLHGVN-----PIILYIARGASIASLSKNDIEFG-- 77
 DB 21 LTIN-----GKAPLVAAVELIARIIVNALPNSIAIKLVDDKPAKALDDATEDVFNKI 75
 QY 78 -----HVVEMLEYAPFL-----SGSEFENACLFVDGFLASRTFLVGHGL--T 118


```

Db 76 TSKPAATFNDGDKQVAKVMIAQKELVTKNFAKISQSLKTLDSQNLRTFLIG-GLKYS 134
Oy 119 IADIAMNSNLGIGORWESLRSKKYONLVMPNSID-----SEKKEALNVAFAVGR 173
Db 135 AADVACMGALNSNGM-CGSIITKKNVDVANSRMTYLLLENDPITFGEHADPLSKSLLETKSA 193
Oy 174 GIGSPAPSLKKEVDSKDPAPAEVDLPQAKGVKCVAFAPESGYLHIGAKAALLNKY 233
Db 194 NVGK-----KKEHTKANF-----EIDLDPADKMGVLTVPFPPEPSYLIHIGAKAALLNOY 243
Oy 234 FAERYQGLIYRPDDTNPSSKSNFENVLLKDIEFLGIKYDAVYTSOYPPKLMEMASL 293
Db 244 FAQAVKGLIIRFDTPNSKEKEEFQDSILEDLIDLGIKGDRITYSYSDYFQEMVDYCVQM 303
Oy 294 IKQKAYIDTTPKEMRKEMRDGIESRCRNNTVENLSLM-KEMWNGTERMGQCCVRKGL 352
Db 304 IKDGAAYCDTTPTEKMRERMDGVASARRDSVEENLRI FTEEMNGTEEGKKNCRVAKI 363
Oy 353 DMODPNKSLRDPVYRCNTDPPHRRVSGSKVYPTDPAFCPPVDALGVTALRSSEYHDR 412
Db 364 DYKALNKLRLRPVYRCNLTTHHRTGTSWKIYPTDFCVPIVDAIEGVTALRTIEYDR 423
Oy 413 NAOYRRIIDQNGLRARVEIYFSRLMVTYLLSKRKLMEFVONKVEDTDRPPTVQIV 472
Db 424 NAOYDMMLQALRLRKVHIWDFARINFRVTLSSKRLQMMVNDKDLGVNMDRPFPTVGR 483
Oy 473 RRGKVELLIQPILOQGSKSNLMEWQKLTINKKIIDPCARTAVLKQORVFTLTN 532
Db 484 RRGVTEELIRFVLSQGSRRVNIEMNLIWAFNKKVIDPAPRTALVNPVKHLEBSE 543
Oy 533 GPEEPFVILIRPHKKFEGAGKATTFANRIMLDVADAANKGSEVTLMGMNMLVKEIK 592
Db 544 APQPKIMKMKKKKPAVGEKKVYIYDIDVDKDAVINDEVTLMGMNVIITK-K 602
Oy 593 VESGVITELVGLLEGSVKTTLKITWADIIEVLPLSLVEFDYLLSKKLEDEDFLD 652
Db 603 NDDG---SMVAKNLNEGDFKTKTKHKLTLWADTKDVPVADVDFHLIKDLLEBDESED 659
Oy 653 NLNCTREIRIPALGDANRNKIKGGIITOLEKRGYRCAPFRSSKPVVLAIPDGR 709
Db 660 FLRTQTEFHTDAIDLNVKDMKIGIIFERRKGYRLDA-LPKGKPYVFFTIIDGK 715

RESULT 3
STEP DROME STANDARD: PRT; 1714 AA.
ID STEP DROME
AC P28668; Q9VCF5;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bifunctional aminoacyl-tRNA synthetase [includes: Glutamyl-tRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase); Prolyl-tRNA synthetase (EC 6.1.1.15) (Proline--tRNA ligase)].
GN AATS-GLUPRO OR CGS394.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peerygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92097547; PubMed=1756734;
RA Cerini C., Kerjan P., Astier M., Gratecos D., Mirande M., Semeriva M.;
RT "A component of the multisynthetase complex is a multifunctional aminoacyl-tRNA synthetase.";
RL EMBL J. 10:4267-4277(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Oregon-R;
RA MEDLINE=97217441; PubMed=9063462;
RA Cerini C., Semeriva M., Gratecos D.;
RT "Evolution of the aminoacyl-tRNA synthetase family and the

```

```

RT organization of the Drosophila glutamyl-prolyl-tRNA synthetase gene.
RT Intron/exon structure of the gene, control of expression of the two
RT mRNAs; selective advantage of the multienzyme complex."
RL Bur. J. Biochem. 244:176-185(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Branton R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Baau A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Buttris K.C., Buzam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunov B.C., Dunn P.,
RA Dudin K.J., Evangelista C.C., Ferraz C., Fertire W.M., Fleischmann W.,
RA Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hootin D., Houston K.A., Howland T.J., Wei M.-H., Ibovcan C.,
RA Jatala M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pectel J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Sytekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers R.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
CC -1- CATALYTIC ACTIVITY: ATP + L-proline + tRNA(Pro) = AMP +
CC -1- CATALYTIC ACTIVITY: ATP + L-proline + tRNA(Pro) = AMP +
CC -1- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS
CC COMPRISED OF A BIFUNCTIONAL GLUTAMYL-PROLYL-TRNA SYNTHETASE, THE
CC MONOSPECIFIC ISOLEUCYL, LEUCYL, GLUTAMINYL, METHIONYL, LYSYL,
CC ARGINYL, AND ASPARYLYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY
CC PROTEINS, P18, P48 AND P43.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CLASS-I
CC AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CLASS-II
CC AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -1- SIMILARITY: Contains 6 WHEP-TRS domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL, M74104; AAA28594.1; -
CC EMBL, U59923; AAC47461.1; -
CC EMBL, AE003745; AAF56211.1; -
CC PIR, S18644; S18644.

```

DR HSPB, P00962; 1GTR.
 DR FLYBae; FBgn005674; Acta-glupro.
 DR InterPro; IPR004526; GltX_arch.
 DR InterPro; IPR000924; GltX_arch.
 DR InterPro; IPR000406; GST_Cterm.
 DR InterPro; IPR004154; GST_Cterm.
 DR InterPro; IPR004499; GST_Cterm.
 DR InterPro; IPR002314; RNA-synt_2b.
 DR InterPro; IPR001412; RNA-synt_1.
 DR InterPro; IPR002316; RNA-synt_pro.
 DR InterPro; IPR006195; RNA_ligase_II.
 DR InterPro; IPR00738; WHEP-TRS.
 DR Pfam; PF00043; GST_C_1.
 DR Pfam; PF03123; HGT_anticodon; 1.
 DR Pfam; PF00749; RNA-synt_1c; 1.
 DR Pfam; PF03950; RNA-synt_1c; 1.
 DR Pfam; PF00587; RNA-synt_2b; 1.
 DR Pfam; PF00458; WHEP-TRS; 6.
 DR PRINTS; PR00987; TRNASYNTHGLU.
 DR PRINTS; PR01046; TRNASYNTHPRO.
 DR TIGRFAMs; TIGR00463; GltX_arch; 1.
 DR TIGRFAMs; TIGR00408; pros_fam_1; 1.
 DR PROSITE; PS00178; AA_TRNA_LIGASE_1; 1.
 DR PROSITE; PS00662; AA_TRNA_LIGASE_1; 1.
 DR PROSITE; PS00762; WHEP-TRS; 6.
 DR AMINOACYL-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;
 KW Multifunctional enzyme; Repeat.
 FT DOMAIN 170 754
 FT BINDING 755 800
 FT DOMAIN 827 872
 FT DOMAIN 901 946
 FT DOMAIN 980 1025
 FT DOMAIN 1055 1100
 FT DOMAIN 1129 1173
 FT DOMAIN 1174 1180
 FT DOMAIN 1207 1714
 FT SITE 209 220
 FT SITE 438 442
 FT BINDING 441 441
 FT CONFLICT 102 106
 FT CONFLICT 233 234
 FT CONFLICT 341 345
 FT CONFLICT 583 583
 FT CONFLICT 692 692
 FT CONFLICT 753 753
 FT CONFLICT 802 802
 FT CONFLICT 873 873
 FT CONFLICT 887 887
 FT CONFLICT 1201 1201
 FT CONFLICT 1461 1461
 FT CONFLICT 1587 1587
 SQ SEQUENCE 1714 AA; 189197 MW; 6FE8C58045E48A8C CRC64;

Query Match 40.1%; Score 1498; DB 1; Length 1714;
 Best Local Similarity 44.2%; Pred. No. 8.1e-96;
 Matches 321; Conservative 126; Mismatches 247; Indels 32; Gaps 13;

QY 1 MEALSPKSDPPIIICAKLVG-LPLTINHSIAAGAPLQFASGESL-HGVNPII 56
 DB 1 MSIKKALNNPPIGLTAHLINGVPEIYMSKEES-LQPPDRRLVCHSNNDVL 57
 QY 57 LYIARGASIASISGNDIEFGHVEWLEVAFTLGSSEFNACLFVDGFLASRTFLVGHG 116
 DB 58 RALAAADPYKLYGTAIERQIDHWLSPLT-CEDISWALSPLTSPPLPVTVLVANK 115
 QY 117 LTIAIAVMSNLAGIGQWESLRKSKYQNLVWRNSIDSEYKELNENVAFAVGRIG 176
 DB 116 LTIDAFALFENM--HSRYEFLAARGKIPQHVQRMWDLTAQ--PLIQKVLQSLPDAVK 170
 QY 177 KSPAPSLKE---KVHSDKQPSAEVDLPAKAVGKCVFAPAPSGYLHGHAKALLNKY 233
 DB 171 RSPQSSKEQTPAKTERKO-EKGFVDLPQAEKGVVAFPPASGYLHGHAKALLNKY 229

QY 234 FAERYGRLTYRPDDTNPESKSENFENLKDIEITLGIKYDAVYTSDYEPKLMEMAE 293
 DB 230 YALVCOGLINRFPDTPAKETVEFENVILLDDQLQKPVFHTSYTFDMLDYCVRL 289
 QY 294 IKQKAVYIDTPKQEMKERMDGIESRCNNVTEENSLMKEMVNGTERGQCCVRGLD 353
 DB 290 IKESKAVYDTPPQOMLERQQRVESANRSVSKNLSLMEWVGSEKGNVCAAKID 349
 QY 354 MODPNKSLRDPVYRCNTDPHRRVGSKYKYPIYDPACPFYDALEGVTHALRSSEYDRN 413
 DB 350 MSSDNGCWDPDTYRCNKEPPEPRGTGYKVPYDFACPIDALENTVHTLRTTEYHNRD 409
 QY 414 AQYRIIDMDGLRVEIYSESRMLNMVYTLISKRLTLPVONKKVEDMTDPFPVQGVIR 473
 DB 410 DQFWFDALKLRPYVMSYSLMTNTVLSKRLTLPVDSGLVDGDDPFPVQGVIR 469
 QY 474 RGLKVEALIOFTLOQAGSKNLNMEMDKLTINKKIIDPVCARTAVLKDQRFVITLNG 533
 DB 470 RGMTEGKERTIAGSSSKSVFPMNDKINAFNKCVIDPIAPRTALKEKRVAVNAGA 529
 QY 534 PEEPFRVRI-LPRHKKFEGAGKATTPANRIMWDYADAANKGEVTLMDGNAIVEI- 591
 DB 530 KVE---RQVSVHDKDSISLGGKTVLGPRIYIDVDAEALKEGNATFINNGNILLIKVN 586
 QY 592 KVEGVTTELVEGLHLEGSVTTTLKITWLA---DIEELVPLSVEEDYILSKKLEDE 648
 DB 587 KDSAGNTSVDAALNENKQFKTLKTLAIVEDDPSPAYVPTFCVYFDNIIISKVLGDE 646
 QY 649 DFLNINPCTREIRIPALGDANMRNIKGEITQLERKGYRCADAFIRSS-----KPVVL 702
 DB 647 DFKPFIHKTRDPVPMIDGDELKCKCKGDIILQORGFVKVAVLPPSGVTNVPSPVL 706
 QY 703 FAIPDG 708
 DB 707 FSIIDG 712

RESULT 4
 SYEP HUMAN STANDARD; PRT; 1440 AA.
 AC P07814;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Bifunctional aminoacyl-tRNA synthetase (includes: Glutamyl-tRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase); Prolyl-tRNA synthetase (EC 6.1.1.15) (Proline--tRNA ligase)).
 GN EPRS OR QPRS OR GLNS OR PARS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91107633; PubMed=1988429;
 RA Felt R., Knippers R.;
 RT "The primary structure of human glutamyl-tRNA synthetase. A highly conserved core, amino acid repeat regions, and homologies with translation elongation factors";
 RL J. Biol. Chem. 266:1448-1455(1991).
 RN [2]
 RP PRELIMINARY SEQUENCE OF 96-887 FROM N.A.
 RC TISSUE=Cervical carcinoma;
 RX MEDLINE=88262551; PubMed=3290852;
 RA Thommes P., Felt R., Schray B., Kunze N., Knippers R.;
 RT "The core region of human glutamyl-tRNA synthetase homologues with the Escherichia coli and yeast enzymes";
 RL Nucleic Acids Res. 16:5391-5406(1988).
 CC -I- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP + diphosphate + L-glutamyl-tRNA(Glu).
 CC -I- CATALYTIC ACTIVITY: ATP + L-proline + tRNA(Pro) = AMP + diphosphate + L-prolyl-tRNA(Pro).
 CC -I- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS

```

CC COMPRISED OF A BIFUNCTIONAL GLUTAMYL-PROXYL-TRNA SYNTHETASE, THE
CC MONOSPECIFIC ISOLEUCYL, LEUCYL, GLUTAMINYL, METHIONYL, LYSYL,
CC ARGINYL, AND ASPARYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY
CC PROTEINS, P18, P48 AND P43.
CC -1 SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO CLASS-I
CC AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -1 SIMILARITY: IN THE C-TERMINAL SECTION, BELONGS TO CLASS-II
CC AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -1 SIMILARITY: Contains 3 WHEP-TRS domains.
CC -1 CAUTION: WAS ORIGINALLY THOUGHT TO BE A GLUTAMINYL-TRNA
CC SYNTHETASE.
CC -1 CAUTION: REF. 2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A NUMBER OF
CC SEQUENCE ERRORS.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its use
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; X54326; CAA38224.1; -.
CC EMBL; X07466; CAA30354.1; ALT_SEQ.
CC PIR; A38663; SYHOUT.
CC PDB; 1FYU; 31-DEC-02.
CC Genew; HGNC:3418; EPRS.
CC MIM; 138295; -.
CC DR GO; GO:0005737; C:cytoplasm; TAS.
CC DR GO; GO:0005625; E:protein complex assembly; TAS.
CC DR GO; GO:0006461; E:protein complex assembly; TAS.
CC DR InterPro; IPR004526; GLX arch.
CC DR InterPro; IPR000924; GLU TRNA-synt_1c.
CC DR InterPro; IPR004046; GST Cterm.
CC DR InterPro; IPR004454; HGTf anticodon.
CC DR InterPro; IPR004499; Prox_fam_I.
CC DR InterPro; IPR002314; tRNA-synt_2b.
CC DR InterPro; IPR001412; tRNA-synt_1.
CC DR InterPro; IPR002316; tRNA-synt_pro.
CC DR InterPro; IPR006195; tRNA_ligase_II.
CC DR InterPro; IPR000738; WHEP-TRS.
CC DR Pfam; PF00043; GST_C_1.
CC DR Pfam; PF03129; HGTf_anticodon; 1.
CC DR Pfam; PF00749; tRNA-synt_1c; 1.
CC DR Pfam; PF03950; tRNA-synt_1c_C; 1.
CC DR Pfam; PF00587; tRNA-synt_2b; 1.
CC DR Pfam; PF00458; WHEP-TRS; 3.
CC DR PRINTS; PRO00987; TRNASYNTHGU.
CC DR PRINTS; PRO1046; TRNASYNTHPRO.
CC DR TIGRFAMs; TIGR00463; glx arch. 1.
CC DR TIGRFAMs; TIGR00408; prox_fam_I; 1.
CC DR PROSITE; PS00178; AA TRNA_LIGASE_I; 1.
CC DR PROSITE; PS00862; AA TRNA_LIGASE_II; 1.
CC DR PROSITE; PS00762; WHEP-TRS; 3.
CC DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Multifunctional enzyme; Repeat; 3D-structure.
CC KW DOMAIN 92 687
CC FT DOMAIN 688 884
CC FT DOMAIN 935 1440
CC FT SITE 132 142
CC FT SITE 360 364
CC FT BINDING 363 363
CC FT DOMAIN 688 733
CC FT DOMAIN 761 806
CC FT DOMAIN 839 884
CC FT DOMAIN 887 919
CC FT SEQUENCE 1440 AA; 163026 MW; C4E185A0AA1C204 CRC64;
CC
CC Query Match 38.5%; Score 1437; DB 1; Length 1440;
CC Best Local Similarity 43.5%; Pred. No. 1,1e-91;
CC Matches 284; Conservative 133; Mismatches 198; Indels 38; Gaps 8;

```

Db	1	MEHTEIDHMLEFSATYKSS	CDS	FT	INELNCLSLRTLYLVNSUSLADLCWATTLKGN	60
Oy	133	ORMSELRSSKXYONLVWFNFSIDSEYKALNEVVA	FVYKGRGI	IGKS	PADSLKEKVD	192
Db	61	AMQBLQKQKAPVHYKRMFGPLEAQ	-QAFQSV	-----	-----	GTRMDVSTTGA 102
Oy	193	PSADE-----	VDLPSAKVGKVCVRAP	PEPSGYLHGHAKAALINKYFAERYOGR	LI	244
Db	103	RVAPEKKODKGVFVELPGAEMGKVTVR	PEPPEASGYLHGHAKAALINQHYQVAFKGLIM			162
Oy	245	RPDTPNEKESNEFENLLKDIETGICYDATYTS	SDVPRKLMEMAE	SLIKGKAI	IDTP	304
Db	163	RPDTPNEKEKEDEKFLVEDVAMHIAHPDQ	FTYISDHEETIMKAEKLIQ	SKAV	AVDT	222
Oy	305	PKEQMRKERMDSIESRCNNVTVEENLSIMKE	VNATERMOCCSVAGK	GLDMODPNKSL	RD	364
Db	223	PABQMKAREGRISKHKHKNPIEKLQWME	BKKSSQFHSCTCLAKIM	SSNGCM	RD	282
Oy	365	VYRRCNTDPHHRVQSKYVYPTPD	FACFPVDAL	EGVTALRSSEYH	RNAQYRILQDNG	424
Db	283	TLVYRCKIDPHRPTGKNVYVPTPD	FACPIVDSIEGVTALR	TEYH	DRDEQFWIT	426
Oy	425	LRRVEIYFESFLNMYTLLSKRKLMPQNK	VEBMTDPR	FPTVOGIVRGLK	VEVLLIQ	484
Db	343	IRKPIWESYSRNLNNTVLSKRKLTFW	FNBEGLVDMDDPR	FPTVRGLRGMT	VEGLKDF	402
Oy	485	ILQOGASNLNLMEMDQKMTINKKIIDP	VCARHTAVLKQRYITFL	TGNGPEEP	VLIDR	544
Db	403	IAQGSSSSVNMEDKIMARKKIKIDP	ARPYALLKKEVIVANV	PAQOE	-MKEVAK	460
Oy	545	HKKEGACKKATTPANRIMLDYADA	AAINKGEVTLMDGNAIVKEI	-KVESG	VITELVG	603
Db	461	HPKQPEVGLKVMVYSPKVFIEG	ADLETSEGEVTFIMNG	INITIKIHKNAD	KIISLDA	520
Oy	604	ELHLEGSKTKTKITMYLADIE	ELVPLSL--VEFY	LISKKLLEBED	LDLNL	PCTRR 661
Db	521	KFNLENQDYKTKTKYMTLAETTHA	PIPIVICTVYEHLLTKP	PLGKDEDEKQYV	NKNSKHE	580
Oy	662	IPALGDANMRNIRKGEITQLE	RKGYRCD-----	AFIRSSKPV	LVFAI	PPG 708
Db	581	ELMIGDPCLKDLKGDIIQ	LORGRGFFIDQ	PIEPSPISCKE	APVLI	IPDG 633

QO	SEQUENCE	554 AA;	63421 MW;	3F36F640BE35D702 CRC64;
FT	BINDING	270	270	ATP (BY SIMILARITY).
FT	SITE	267	271	"KMSKS" REGION.
FT	SITE	33	43	BY SIMILARITY.
KW	INIT_MET	0	0	HIGH REGION.
KM	Complete proteome.			
DV	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;			
DR	PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.			
DR	TIGRFAMs; TIGR00440; GlnS; 1.			
DR	PRINTS; PR00987; TRNAsYNTHGU.			
DR	Pfam; PF03950; tRNA-synt_1c; C; 1.			
DR	InterPro; IPR001412; tRNA-synt_1.			
DR	InterPro; IPR000942; Glu_tRNA-synt_1c.			
DR	GlnS; GlnS; GlnS.			
DR	HAMAP; MF_00126; -; 1.			
DR	EMBL; AEO16841; AACG9799.1; -.			
CC	-			
CC	or send an email to license@lsb-sib.ch).			
CC	-			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@lsb-sib.ch).			
CC	-			
CC	-1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.			
CC	-1- SUBUNIT: Monomer (By similarity).			
CC	-1- PHOSPHATE: diphosphate + L-glutaminyl-tRNA(Gln).			
CC	-1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP + diphosphate + L-glutaminyl-tRNA(Gln).			
RL	J. Bacteriol. 185:2330-2337(2003)			
RA	"Comparative genomics of Salmonella enterica serovar Typh strains Ty2 and Cyt8."			
RA	Butland V., Kodymiani V., Schwartz D.C., Blatner F.R.;			
RX	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,			
RC	MEDLINE=22553367; PubMed=12644504;			
RP	SEQUENCE FROM N.A. ATCC 700931;			
[2]				

Query Match	Similarity	20.3%	Score 757.5	DB 1	length 554
Best Local	Similarity	35.1%	Pred. No. 4,6e-45		
Matches	189	Conservative	95	Mismatches	201
				Indels	53
				Gaps	15
QY	199	DLPGAKVKCVCFAPBPSPGYLHHGAKAALINKYFAERYGRLIVRPDDTNPSKESNEF	258		
Db	18	DLASGKHVTVHTRFPPEPENGYLHHGAKSICLNFGIADYQGCNLRPDDTVPEIDLEY	77		
QY	259	VENLLKDIETGIKKYA-VITYSDVFPKLMWASBLINQKRAYIDDTKEQWRKER----	313		
Db	78	VDSIKNDVEMWGFPHSGDIRYSSDFDQHHAYAVELINKGLAYVELPPEQIREYGTLL	137		
QY	314	MDGIESRCRNNTVEENLSLTKMEM-NGTERGMQCCVRGLDMQDNKSLRPDYVRCNTD	372		
Db	138	APGKNSPPRDRSVEENLALFEKMRIGGPEEG-KACLRPAKIDWASFVIMRDPVLRIKPA	196		
QY	373	PHHRVGSKKYKPYTYDFACPFVDALGVTTHLRSSSEYHNRNAAQYRRILODMGLR-RVEIY	431		
Db	197	EHHQGTGTCWCIYPMWDFTHCISDALEGITHSCLTEPFDNRRLXYMWLVNDNITIPHAPROX	256		
QY	432	EFSPRLNMYVYLLSKRKLMPYONKRYEVDTPRPTVGIYVRGKLVEALLIOTLQOQAS	491		
Db	257	EFSPRLNLEYTVWSKSKLNLVTDKRVBEGDDPRMPTISGLRRRGYTAASIREFCRIGVT	316		
QY	492	KNLNLMEWDKLTWINKKIIDVPCARHTAVLKQDRIYFTLNGPE-EFPVRLIPRHKKEG	550		
Db	317	KQDNITIEWASLESCEIREDLNENAPAMAVIDPVKLV-IENYQGESEMYVTPMNPKNKE	374		
QY	551	AGKATTTANRMLDYAD-AAALNK-----GEEVTLMDKNA-IYKIKYESGVITEL	601		
Db	375	MGSRVPPSGEWTIDRADPREEANQYKRLVNGKVRIR---NAVVIKAEERKXA----	427		

[illegible]

	RESULT 6
SYQ_SALTY	
ID_SYQ_SALTY	STANDARD; PRT; 554 AA.
AC	Q8ZQX5;
DT	28-FEB-2003 (Rel. 41, Created)
DT	28-FEB-2003 (Rel. 41, Last sequence update)
DT	28-FEB-2003 (Rel. 41, Last annotation update)
DE	Glutaminyl--tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase) (GlnRS).
DE	
GN	GINS OR STM0686.
OS	<i>Salmonella typhimurium</i> .
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC	Enterobacteriaceae; <i>Salmonella</i> .
OX	NCBI_TaxID=602;
OX	1511

RP SEQUENCE FROM N.A.
RC STRAIN=L172 / SGCSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L.D., Porwollik S., Ali T., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoeckert T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
RT L172".
RL Nature 413:852-856(2001).
CC -I- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
CC diphosphate + L-glutaminyl-tRNA(Gln).
CC -I- SUBUNIT: Monomer (By similarity).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announce/>
CC or send an email to license@ebi.ac.uk).

DR	EMBL; AE008727; AAL19630.1; -. StyGene, SG7777; glns.				
DR	HAVAP; MF_001261; -. InterPro; IPR004514; Glns.				
DR	InterPro; IPR000924; Gly_CrNA-synt_1c.				
DR	InterPro; IPR001412; CrNA-synt_1.				
DR	Pfam; PF00749; CrNA-synt_1c; 1.				
DR	Pfam; PF03950; CrNA-synt_1c; 1.				
DR	PRINTS; PR00987; TRNASYNTHGU.				
DR	TIGRFAMs; TIGR00440; Glns; 1.				
DR	PROSITE; PS00178; AA TRNA LIASE I; 1.				
KW	Aminacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding				
KW	Complete proteome.				
FT	INT_MET	0			
FT	SITE	33	43		BY SIMILARITY.
FT	SITE	267	271		"HIGH" REGION.
FT	BINDING	270	270		"KMSKS" REGION.
FT	SEQUENCE	554 AA;	63406 MW;	8P47CEA5B65BEAD CRC64;	ATP (BY SIMILARITY).

Query Match	20.2%	Score 755.5;	DB 1;	Length 554;
Best Local Similarity	35.3%;	Pred. No. 6.4e-45;		
Matches 190;	Conservative 93;	Mismatches 202;	Indels 53;	Gaps 15;

QY 199 DLPGAKVGVKVCRRAPPEPSGTHIGHAKAALINKYFAERYGRLIVRPDITNPSKESNEF 258
 DB 18 DLASGKHHTVTRFPPEPNBYLHIGAKSICLNFIADYQOQCLRFDDITPVXEDIEY 77
 QY 259 VENLLKDIETIGIKKDA-VTVSDVPEPLKEMASBLIKQKAYIDDTREKQKRKR----- 313
 DB 78 VDSINNDVEMWGFPHSGDRIYSSDYFDQHHAYAVELINKGLAYDELPEQIREYRGTLT 137
 QY 314 MDGIESRCNNVTVENLSIKMEM-VNGTERGMOCCVGRGLMDQDNKSLRDPVYRCNTD 372
 DB 138 APGKNSPRDRSVEENLALFEKMRGGEPEG-KALIRAKIDMASPIYMRDPVLIRIFA 196
 QY 373 PPHRVGSXYKYPTYPDACPFDVALEGVTHALRSSEYHDNRNAQYVRIIDMGLR-RVEIY 431
 DB 197 EHHQGNKMCITYPMWDFTHCISDALEGITHSLCTLEFDQNRRLYMWVLDNITIPVHPROY 256
 QY 432 EFSRLNMYTLLSKKLMFVNQKVEDWTPRPFTVQGIYRGLKVEALLOFIQOGAS 491
 DB 257 EFSRLNMYTLLSKKLMFVNQKVEDWTPRPFTVQGIYRGLKVEALLOFIQOGAS 491
 QY 492 KNLNLMEMDKLWTINKKIIDPVCARHTAVLKQDQVIFLTNGPE-EPPVRIIPRHKFEFG 550
 DB 317 KQDNTIEWASLESCTREDLNENAPAMAVIDPVKLV--IENYPOGESEMTVMPHPNKE 374
 QY 551 AGKKAATTPANRIWLDYAD-AAAIK-----GEEVTLMDWGN-IVKEIKVESGITEL 601
 DB 375 MGSREVPFSGEIMWRADPREANQYKRLVWGKEVRLR--NAVIAERENKXA----- 427
 QY 602 VGEHLHESGSKT-----TKLK--ITWLADIEELVPLSLVEFDYLISK 642
 DB 428 -----ENITITTFCTYDADTLKSDPADGRKVKGYIHWSAHAL-PIEIRLYDRHFGVP 480
 QY 643 KLEBEDFLDNLNPGCTREIRPALGDANMNRKGIILDERKGYRCAPFRSSKPV 700
 DB 481 NPGAAEDFLSVINPESLVIKQGYGEPSLKAAVAGAFQFERGYCLDSRYATADKLV 538
 RESULT 7
 SYO_ECOLI
 ID SYO_ECOLI STANDARD; PRT; 553 AA.
 AC P00962; Q59403;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine-tRNA ligase)
 GN (GlnRS).
 OS GlnRS OR B0680.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 CX NCBI_TaxID=562;
 RN (1)
 RN SEQUENCE FROM N.A.
 RP MEDLINE=89113343; PubMed=2464170;
 RA Uemura H., Conley J., Yamao F., Rogers J., Soell D.G.;
 RT "Escherichia coli glutaminyl-tRNA synthetase: a single amino acid
 RT replacement relaxes tRNA specificity.";
 RL Protein Seq. Data Anal. 1:479-485(1988).
 RN (2)
 RN SEQUENCE FROM N.A.
 RP STRAIN=K12;
 RC MEDLINE=83007236; PubMed=6288695;
 RA Yamao F., Inokuchi H., Cheung A., Ozeki H., Soell D.G.;
 RT "Sequence of the glnS gene.";
 RL J. Biol. Chem. 257:11639-11643(1982).
 RN (3)
 RN SEQUENCE FROM N.A.
 RP MEDLINE=85051900; PubMed=6389180;
 RA Hoben P., Uemura H., Yamao F., Cheung A., Swanson R., Summer-Smith M.,
 RA Soell D.;
 RT "Misaminoacylation by glutaminyl-tRNA synthetase: relaxed specificity
 RT in wild-type and mutant enzymes."

RL Ped. Proc. 43:2972-2976(1984).
 RN (4)
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatcher F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirtpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN (5)
 RN SEQUENCE FROM N.A.
 RP STRAIN=K12;
 RC MEDLINE=97061202; PubMed=6905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Takami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN (6)
 RN SEQUENCE OF 1-12;
 RP STRAIN=K12 / EMG2;
 RX MEDLINE=97443975; PubMed=9298646;
 RA Link A.J., Robinson K., Church G.M.;
 RT "Comparing the predicted and observed properties of proteins encoded
 RT in the genome of Escherichia coli K-12.";
 RL Electrophoresis 18:1259-1313(1997).
 RN (7)
 RN SEQUENCE OF 1-5, AND CHARACTERIZATION.
 RC STRAIN=K12;
 RX MEDLINE=83007237; PubMed=6749844;
 RA Hoben P., Royal N., Cheung A., Yamao F., Biemann K., Soell D.;
 RT "Escherichia coli glutaminyl-tRNA synthetase. II. Characterization of
 RT the glnS gene product.";
 RL J. Biol. Chem. 257:11644-11650(1982).
 RN (8)
 RN X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RP MEDLINE=90069584; PubMed=2479982;
 RX Rould M.A., Perona J.J., Soell D., Steitz T.A.;
 RT "Structure of E. coli glutaminyl-tRNA synthetase complexed with
 RT tRNA(Gln) and ATP at 2.8-A resolution.";
 RL Science 246:1135-1142(1989).
 RN (9)
 RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RP MEDLINE=91312443; PubMed=1857417;
 RX Rould M.A., Perona J.J., Steitz T.A.;
 RT "Structural basis of anticodon loop recognition by glutaminyl-tRNA
 RT synthetase.";
 RL Nature 352:213-218(1991).
 RN (10)
 RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RP MEDLINE=96230743; PubMed=9562563;
 RX Rath V.L., Sullivan L.F., Beijer B., Sproat B.S., Steitz T.A.;
 RT "How glutaminyl-tRNA synthetase selects glutamine.";
 RL Structure 6:439-449(1998).
 RN (11)
 RN X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RP MEDLINE=2031876; PubMed=10860750;
 RA Shearin L.D., Bullock T.L., Newberry K.J., Lipman R.S., Hou Y.M.,
 RA Beijer B., Sproat B.S., Perona J.J.;
 RT "Influence of transfer RNA tertiary structure on aminoacylation
 RT efficiency by glutaminyl and cysteinyl-tRNA synthetases.";
 RL J. Mol. Biol. 299:431-446(2000).
 CC -I- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
 CC -I- diphosphate + L-glutaminyl-tRNA(Gln).
 CC -I- SUBUNIT: Monomer.
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -I- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G., III, Redford P., Roesech P.,
 RA Raesho D., Buckles B.L., Lion S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
 CC diphosphate + L-glutamyl-tRNA(Gln).
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to class-I aminacyl-tRNA synthetase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, AE016751; AAN79239.1; -.
 DR HAMAP, MF_00126; -; 1.
 DR Pfam; PF00749; tRNA-synt_1c; 1.
 DR Pfam; PF03950; tRNA-synt_1c C; 1.
 DR TIGRFAMs; TIGR00440; glnS; 1.
 DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
 KM Aminocacyl-tRNA synthetase; protein biosynthesis; ligase; ATP-binding;
 KM Complete proteome.
 FT INIT_MET 0 BY SIMILARITY.
 FT SITE 33 43 "HIGH" REGION.
 FT SITE 267 271 "KMSKS" REGION.
 FT BINDING 270 270 ATP (By SIMILARITY).
 FT SEQUENCE 553 AA; 63362 MW; FE89470BA128D8F9 CRC64;
 SO
 Query Match 20.1%; Score 750; DB 1; Length 553;
 Best Local Similarity 35.0%; Pred. No. 1.5e-44;
 Matches 188; Conservative 92; Mismatches 205; Indels 52; Gaps 14;
 Oy 199 DLPAKVKVCVPAPEPSGYLHGAALANKYFAERYOQLIVRPDTPNPKSENEF 258
 Db 18 DLASGKHTTVTRPPEPNGLYHIGAKSICLNFGLAQDYKQCNLRFDDTNPVKEIDY 77
 Oy 259 VENLKDLETGIGKDA-VYTSDFPFLGMEASLIKQKAYIDDTREKQMKER---- 313
 Db 78 VDSIKNDVEMWGFHMSGNVRYSDFDQLHAYAILINKGLAYVDELPEQIREYRGILT 137
 Oy 314 MDGIESCRNNTVEENLSLAKEM-VNGTERGQCVCVRGLMDQDNKSLRPVRYRCNTD 372
 Db 138 QPKKSPFRDSSVEENLALPEKMTGSGEBG-KACLRAKIDIASPFYWRDVLRYIFA 196
 Oy 373 PPHRVGKYKVPYTPDFACPVDALGVTALRSSEYHNRNAQYRIILQDMGLR-RVEIY 431
 Db 197 EHHQGNKWCICYPMWDFHCISDALGEGITHSLCTLEFDQNRRLYQWVLDNITIPHPROY 256
 Oy 432 EFSRLNMYTILSKRLMFPONKKEVDWTDPRPTVOGIYRGLKVEALIOFTIIQOGAS 491
 Db 257 EFSRLNMYTILSKRLMFPONKKEVDWTDPRPTVOGIYRGLKVEALIOFTIIQOGAS 491
 Oy 492 KNLNIMEDYKLTWINKKTIIDPYCAHRTAVLKDQQRVIFLTNGSPREPPFRLIPRHKKFEGA 551
 Db 317 KODNTIEVASLSECTREDLNEAPAPAAVADPYKVIENYQEGE--VWTMHNHKKPEM 374
 Oy 552 GKKAITPAMRIWLDYAD-AAAIK-----GEEVTLMDWGA-IVKEIKYESSGIVTELV 602
 Db 375 GSRQVPFGEGLWIDRADFREANQKQKRLVIGKEVRLR---NAVYIKAREPKDA----- 426

Oy 603 GELHEGSEVKT-----TKLK--ITWLDIEELVPLSVERDYLSKKK 643
 Db 427 -----EGNITITIFCYADDTLSKOPADRKTKGVHWSAHAAL-PVEIRLVDRFSPVFN 480
 Oy 644 LEEDEFLDNLNICTRREIPALGDNMNIRKEIIGLERGYKCDAPFIRSSRPV 700
 Db 481 PGAADDFLSVINPESLVIKQGFAPESPGLKDAVAKRPFQEREGYFCLDSRHSTAEKPV 537
 RESULT 9
 SYQ_ECO57 STANDARD; PRT; 553 AA.
 AC O8X98;
 DT 28-FEB-2003 (Rel. 41, last sequence update)
 DT 28-FEB-2003 (Rel. 41, last annotation update)
 DE Glutamyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
 DE (GlnRS).
 GN GlnS OR 20827 OR EGS0710.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533 (2001).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22 (2001).
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
 CC diphosphate + L-glutamyl-tRNA(Gln).
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to class-I aminacyl-tRNA synthetase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, AE005246; AAG55002.1; -.
 DR EMBL, AP002552; BAB34133.1; -.
 DR PIR, F85567; F85567.
 DR PIR, F90717; F90717.
 DR HAMAP, MF_00126; -; 1.
 DR InterPro; IPR004514; GlnS.
 DR InterPro; IPR000924; Gln_sRNA-synt_1c.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF00749; tRNA-synt_1c; 1.
 DR Pfam; PF03950; tRNA-synt_1c C; 1.
 DR PRINTS; PR00987; TRNASYNTGTLU.
 DR TIGRFAMs; TIGR00440; glnS; 1.

DR PROSITE: PS00178; AA TRNA LIGASE I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KM Complete proteome.
 FT INIT MET 0 BY SIMILARITY.
 FT SITE 33 43 "HIGH" REGION.
 FT SITE 267 271 "KMSKS" REGION.
 FT BINDING 270 270 ATP (BY SIMILARITY).
 SQ SEQUENCE 553 AA: 63378 MW: 44951 DEB8742567 CRC64;
 Query Match 20.0%; Score 747; DB 1; Length 553;
 Best Local Similarity 34.8%; Pred. No. 2.5e-44;
 Matches 187; Conservative 93; Mismatches 205; Indels 52; Gaps 14;
 QY 199 DLPGAQKVCVAPRPPSGYLIGHAKALLNKYFAERYGRILVRPDDTNPSEKSENF 258
 DB DLASGKHTVHTRPPEPENGYLIGHAKSICLNFGIADYQGCNLRPDDNPVEDVEF 77
 QY 259 VENLKDIETLTGIKDA-VVYTSDFPKLMEASLTKOGKAYIDDTPEKQMKRER---- 313
 DB VDSIKNDVEMWLGFRKSGNVRYSDFDQLHAYALIELINKGLAYDELTPREQIREYRGLT 137
 QY 314 MDGIESRCRNTTVEENISLWKEM-VNGTERGQCVCVRGLDMQDNKSLRDPVYRCNTD 372
 DB QPGKSPYRDRSVEENLALFEKMRGTGPEEG-KACLRAKIDWASPFIWRDPLVRIKEFA 196
 QY 373 PHHVGSKYKVPYTFDACPFDALGVTTHALRSSEYHDNRNAYRIIDQMGIR-RVEIY 431
 DB 197 EHHGTGKMCYIPWDFTHCISDALEGITHSLCTLEFQDNRLXWVLDNITIPVHPROY 256
 QY 432 EFSRLNMYTLLSKRKLMPYONKVEEDTDPREFPTVOGIVRGKLVKALIOFLIOGAS 491
 DB EFSRLNMYTLLSKRKLMPYONKVEEDTDPREFPTVOGIVRGKLVKALIOFLIOGAS 491
 QY 492 KNLNLMEMDKLMTINKKIIDVCAHNTAVLKQDQVIFLTNGPEEPFVRLIPRHKKFECA 551
 DB 317 KQDNTIEMASLESICREDLNENAPAMVIDPVKLVINYGEGE--VWTMNNHNKPEM 374
 QY 552 GKKAATTPANRIWLVDYD-AAALNK-----GEEVTLMDMGN-IVKEIKVESGYITELV 602
 DB 375 GSRQVPFGEIWIWDSDFREBANKQYKRLVIGKEVRL---NAVYIKAEVKEKA----- 426
 QY 603 GELHLEGSVKT-----TKLK--ITWADIEELPVLISVEDVYLISKKK 643
 DB 427 -----EGVITTFCTYDADTLISKDPAQGRKVGVIHWASAHAL-PVEIRLYDLRFSVBN 480
 QY 644 LEEBDDPLDNLNICTRREIRIPALGDANMRNIKGEIIOERKGYRCDAFPIRSSKPV 700
 DB 481 PGAADDFLSVINPESLVIKQGAEPSTLKDVAAGKAFQFERREGVFCLDGRHSTAEKPV 537
 RESULT 10
 SYQ_YERPE STANDARD; PRT; 555 AA.
 ID_O82DD9;
 AC 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
 GN GINS OR YPO2630 OR Y1205.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Tlball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebatina M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooke K., Cerdeno-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltham T., Hamlin N., Holtroyd S., Jagers K., Karlyshev A.V.,

RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.,
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
 RL Nature 413:523-527(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIMS / Biovar Mediaevalis;
 RX MEDLINE=22137863; PubMed=12142430;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Fehrerston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
 RA Perry R.D.;
 RT "Genome sequence of Yersinia pestis KIM.";
 RL J. Bacteriol. 184:4601-4611(2002).
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
 CC -1- dihydrophate + L-glutamyl-tRNA(Gln).
 CC -1- SUBUNIT: Monomer (by similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL; AJ14153; CAC92873.1; "-"
 DR EMBL; AE013724; AAM84782.1; ALT_INIT.
 DR PIR; AB0321; AB0321.
 DR HAMAP; MF_00126; "-".
 DR InterPro; IPR004514; GlnS.
 DR InterPro; IPR000924; Gln tRNA-synt_1c.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF00749; tRNA-synt_1c; 1.
 DR Pfam; PF03950; tRNA-synt_1c; 1.
 DR PRINTS; PR00987; TRNASYNTHGL.
 DR TIGRFAMs; TIGR00440; glnS; 1.
 DR PROSITE; PS00178; AA TRNA LIGASE I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KM Complete proteome.
 FT SITE 34 44 "HIGH" REGION.
 FT SITE 268 272 "KMSKS" REGION.
 FT BINDING 271 271 ATP (BY SIMILARITY).
 SQ SEQUENCE 555 AA: 63750 MW: 40368 DEB56C0B5 CRC64;
 Query Match 19.9%; Score 743.5; DB 1; Length 555;
 Best Local Similarity 36.8%; Pred. No. 4.4e-44;
 Matches 196; Conservative 80; Mismatches 214; Indels 43; Gaps 15;
 QY 199 DLPGAQKVCVAPRPPSGYLIGHAKALLNKYFAERYGRILVRPDDTNPSEKSENF 258
 DB DLASGKHTVHTRPPEPENGYLIGHAKSICLNFGIADYQGCNLRPDDNPVEDVEF 78
 QY 259 VENLKDIETLTGIKDA-VVYTSDFPKLMEASLTKOGKAYIDDTPEKQMKRER---- 313
 DB VDSIKNDVEMWLGFRKSGNVRYSDFDQLHAYALIELINKGLAYDELTPREQIREYRGLT 138
 QY 314 MDGIESRCRNTTVEENISLWKEM-VNGTERGQCVCVRGLDMQDNKSLRDPVYRCNTD 373
 DB APGKSPYRDRSVEENLALFEKMRGTGPEEG-KACLRAKIDWASPFIWRDPLVRIKEFA 198
 QY 374 HHRVSKYKVPYTFDACPFDALGVTTHALRSSEYHDNRNAYRIIDQMGIR-RVEIY 432
 DB HHGSKNKCIYPMWDFTHCISDALEGITHSLCTLEFQDNRLXWVLDNITIPVHPROY 258
 QY 433 FSRILNMYTLLSKRKLMPYONKVEEDTDPREFPTVOGIVRGKLVKALIOFLIOGASK 492
 DB 259 FSRILNMYTLLSKRKLMPYONKVEEDTDPREFPTVOGIVRGKLVKALIOFLIOGASK 492
 QY 493 NNLNLMEMDKLMTINKKIIDVCAHNTAVLKQDQVIFLTNGPEEPFVRLIPRHKKFECA 552


```

Db      319 QDNVNMWMSLESCIDDLNEHAPRAMAVLDPIKVIYENRAGEEWLT--MPNHPNNPEWG 376
Oy      553 KKATFANRIMLDYAD-AAAIK-----GEEVTLMDMGNA-IYKEIYVE---SGVILE 600
Db      377 SRQVFDSDSTYIDRADFREERANKQYKRLVLSKEVRLR--NAYVIAKEVDEADAGNVTT 433
Oy      601 LVGELHLR-----GSVTKTKLKITWLAIDIEELVPLSLVEFDYLLSKKLEDEDFDLNL 654
Db      434 LYCSYDAETLNKDPADGRKVKGVIMHVS-VAAHALPAEIRLRYLRLNVPNPAADPFLSTI 492
Oy      655 NP---CTRREI--PALGDANMENIKRGETIQLERKGYRCDAPIRRSKPVVL 702
Db      493 NPESLIVRQGFEPESLADA-----VSDKTYQFEREGYFCADS---RYSRPAAL 537

RESULT 11
STQ_VIBPA
ID      SYQ_VIBPA  STANDARD;  PRT;  556 AA.
AC      Q87G4;
DT      15-SEP-2003 (Rel. 42, Created)
DT      15-SEP-2003 (Rel. 42, Last sequence update)
DE      15-SEP-2003 (Rel. 42, Last annotation update)
DE      Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
GN      (GlnRS).
OS      GlnRS OR VP0832.
OC      Vibrio parahaemolyticus.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OX      Vibrionaceae; Vibrrio.
OX      NCBI_TaxID=670;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=RMD 2210633 / Serotype O3:K6;
RX      MEDLINE=22508454; PubMed=12620739;
RA      Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA      Iijima Y., Nishijima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA      Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT      "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT      distinct from that of V. cholerae.";
RL      Lancet 361:743-749(2003).
CC      -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
CC      diphosphate + L-glutaminyl-tRNA(Gln).
CC      -1- SUBUNIT: Monomer (By similarity).
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC      -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      HMAP; AP005075; BACS9095.1; -.
DR      HAMAP; MF_00126; -. 1.
DR      PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW      Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW      Complete proteome.
FT      SITE      34 "HIGH" REGION.
FT      SITE      268 "RMSKS" REGION.
FT      BINDING  271 272 ATP (By similarity).
SO      SEQUENCE  556 AA; 63969 MW; 355B526332146B44 CRC64;

Query Match      19.7%; Score 734.5; DB 1; Length 556;
Best Local Similarity 34.8%; Pred. No. 1.9e-43;
Matches 184; Conservative 91; Mismatches 201; Indels 53; Gaps 12;

Oy      199 DLPAKVGKVCYRFPSPSGYHIGAKAALINKTFAEYRQRLIVRPDTPNPSKESNEF 258
Db      19  DLADKHTSVHTRFPPEPGYLIHGAIXSICLNFGLADPYQOCNLRFPDTPKEDIEY 78
Oy      259 VENILKDIETIGIKYDA-VTYTSDYFPKLMEWAESLIGKRAYIDTDPREQRKER----- 313

```

```

Db      79 VESIKKDVNVLGFEDWGCVCSSNYFDLKYEAIELINKGLAYDELSPEQIREYRGTLK 138
Oy      314 MDGIESRCRRNTVEENLSLAKEMVNGTERGMOCCRGGLDMODPKSLRDPYVYCNTPD 373
Db      139 EPGKSPYRDSVEENLALPEKTRRGEFEBGACIRAKIDGSSFWMDPVLVYRFPAT 198
Oy      374 HHRVGSKYKVPPTYYPACPFVDALRGVTHALRSSSEYHRNAQYRILIDMGIR-VWEIYE 432
Db      139 HQTGDKKICITPMYDFTHCISDALBGIHSTLTLEFNNRRLYDVLNLTITDGCPRHOYE 258
Oy      433 FSRLLMVTLLSKRKLMLFVONKVEEDWTPRPFVGIIVRGKLEVALIQFILOQASK 492
Db      259 FSRLLMLEYTVMSKRKLMLVTEKLVNGWDDPRMPVSGLRGRGFPASIRECKRIGVTK 318
Oy      493 NLNLMEMKLTINKKIIDPVCARITVLKQORYL-----FTLTNPEEPFARI 541
Db      319 QENMIEFSSLESCIRDDLNENAPRAMAVLDPIKVIYENRAGEVAVENLTLANHPNPEM-- 376
Oy      542 LPRHKKFPGAGKATTFANRIMLDYAD-AAAIK-----GEEVTLMDMGNAIYKEIKV 593
Db      377 -----GERVPTFREWIEREDFREERANKYKRLVLSKEVRLR--GAYVIAKEV 424
Oy      594 E---SGVIT-----ELVGEHLDEGSVTKTKLKITWLAIDIEELVPLSLVEFDYLLSKK 642
Db      425 EKDAEGNITTTTYCTYDPEFTLGNPADG--RKVKGVIMHVS-ADKALPAEIRLYRDLFTVP 481
Oy      643 KLEDEDFDLNINPCTRREIPALGDANMENIKRGETIQLERKGYRCDA 691
Db      482 NPAAADEPASTINTDSLVLVINGFVPSLASAEAGOGYQFERMGYFCADS 530

RESULT 12
STQ_PASMU
ID      SYQ_PASMU  STANDARD;  PRT;  559 AA.
AC      P57847;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DE      16-OCT-2001 (Rel. 40, Last annotation update)
DE      28-FEB-2003 (Rel. 41, Last sequence update)
DE      Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
GN      (GlnRS).
OS      GlnRS OR PM0528.
OC      Pasteurella multocida.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC      Pasteurellaceae; Pasteurella.
OX      NCBI_TaxID=747;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=PM70;
RX      MEDLINE=21145866; PubMed=11248100;
RA      May B.J., Zhang Q., Li L.U., Faustian M.L., Whitlam T.S., Kapur V.;
RT      "Complete genomic sequence of Pasteurella multocida Pa70.";
RT      Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC      -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
CC      diphosphate + L-glutaminyl-tRNA(Gln).
CC      -1- SUBUNIT: Monomer (By similarity).
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC      -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE006088; AAK02612.1; -.
DR      HSSP; P00962; IGTR.
DR      HAMAP; MF_00126; -. 1.
DR      InterPro; IPR004514; GlnS.
DR      InterPro; IPR000924; Gln tRNA-synt_1c.
DR      InterPro; IPR001412; tRNA-synt_1.

```

DR Pfam; PF00749; tRNA-synt 1c; 1.
 DR Pfam; PF03950; tRNA-synt 1c; 1.
 DR PRINTS; PRO0987; TRNASYNTHGU.
 DR TIGRFAMs; TIGR00440; glns; 1.
 DR PROSITE; PS00178; AA TRNA_LIGASE_I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT SITE 44 54 "HIGH" REGION.
 FT SITE 279 283 "KMSKS" REGION.
 FT BINDING 282 282 ATP (BY SIMILARITY).
 SQ SEQUENCE 559 AA; 64309 MW; C3DBC90B65F071D1 CRC64;
 Query Match 19.4%; Score 725; DB 1; Length 559;
 Best Local Similarity 33.5%; Pred. No. 8.6e-43;
 Matches 189; Conservative 101; Mismatches 223; Indels 52; Gaps 15;
 QY 162 LNEVVAEVGKRGKIGKSPAPSLKEKVDHSDKSPADEVLPGAKVGVCPAPSPGGLH 221
 DB 5 VENVVA-----ENTKRPNTNPIRQID-----DLASGKHGVOTRPPPENGYLH 51
 QY 222 IGHAKAALLNKYFAERYQRLIVRPDTPNPSKESNEFVENLLKDIETLGIKYDA-VITYS 280
 DB 52 IGHAKSICLNFAGIADYQGLCNLRDPTNPVEDYVDSIKQVEMLGPKWEGSPRYAS 111
 QY 281 DYFPLMAESLIQGAAYIDDTKQMKRER---MDGISRCNNTVENVLSMKEM 336
 DB 112 DYFQOLYGAIELKEGLAYVELSPEMREYRGTLTEPGKNSPYRDSIEENLAFBKM 171
 QY 337 VNGTERGQCCVRGLDMQDPNKSRLDPVYVRCNTDPHHRVSKYKVPYTFDFACPFYDA 396
 DB 172 KNGELAEKACLRKIDMASPTVWRDVIYRIKATHQIGDKKCIIPMDFTHCISDA 231
 QY 397 LEGVTHALRSSEYHDRAQYRIIDQMGRLR--VEIYFSRLNVTYLLSKRKLMPYQN 454
 DB 232 IERTHSCTLEFQONRLYDMVLNENISARLPHQYEFSLNLESTLTSKRLQLVNE 291
 QY 455 KKVEDWTPRPPTVQGIYRGKLVKVALIQFLIQGASAKLNMEDKMTINKIIDIYVC 514
 DB 292 GIVGWMNDPRMPTISGLRRRGTPASLSRPFRCIGIVTQDNMNVESALAEACIREDLNNA 351
 QY 515 ARHTAVLDQVIFLITNGPEEPFRILPRHKKPFAGAKKATTFANRIMLQVAD-AAAIN 573
 DB 352 PRMAAVIPKLIVENFSGKE---MLTAPNPNRDELGVRELFPRELYIDQADREAN 408
 QY 574 K-----GEEVTLMDGNA-IVKEIKYSGV---IT-----ELVGEHLGSGVXTT 614
 DB 409 KQYKRLVYGKEVRLR---NAVYIKAREVKEAEEINTTVYCTYDPTLGKQPADG--RKV 463
 QY 615 KLTQTLADIEELVLSLVEFDYLSKKKLEDEDFDLNLPCTRRRIIPALGDANMRKIK 674
 DB 464 KGVLOWVS-ABDYLPFAERFOYGRLTVANPGAAEDIHQVLNPDLSVIKQGVVEKSIANAQ 522
 QY 675 RGEITLERKGYRGCDAPFIRSKP 699
 DB 523 PEKAYQFEREGYTCADS---KDSRP 544
 RESULT 13
 SYQ_VIBCH STANDARD; PRT; 556 AA.
 AC Q9KTA6;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DT 28-FEB-2003 (Rel. 41; Last annotation update)
 DE Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
 GN GINS OR VC0997.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N. A.

RC STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10955301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Hatt D.H., Hickey E.K., Peterson J.D., Umeyam U.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,
 RA McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae.";
 RL Nature 406:477-483(2000).
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
 CC -1- SUBUNIT: Monomer (by similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE004181; AAF94158.1; -
 CC PIR; H82254; H82254.
 CC HSSP; P00962; 1GTR.
 CC TIGR; VC0997; -
 CC HAMAP; MF_00126; -; 1.
 CC InterPro; IPR004514; GINS.
 CC InterPro; IPR000924; Gln tRNA-synt 1c.
 CC InterPro; IPR001412; tRNA-synt 1.
 DR Pfam; PF00749; tRNA-synt 1c; 1.
 DR Pfam; PF03950; tRNA-synt 1c; 1.
 DR PRINTS; PRO0987; TRNASYNTHGU.
 DR TIGRFAMs; TIGR00440; glns; 1.
 DR PROSITE; PS00178; AA TRNA_LIGASE_I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT SITE 34 44 "HIGH" REGION.
 FT SITE 268 272 "KMSKS" REGION.
 FT BINDING 271 271 ATP (BY SIMILARITY).
 SQ SEQUENCE 556 AA; 64089 MW; 8108C036F0760F8 CRC64;
 Query Match 19.3%; Score 721; DB 1; Length 556;
 Best Local Similarity 34.3%; Pred. No. 1.6e-42;
 Matches 185; Conservative 91; Mismatches 208; Indels 56; Gaps 13;
 QY 199 DLPQAVGKVCVAPAPSPGGLHIGHAKAALLNKYFAERYQRLIVRPDTPNPSKESNEF 258
 DB 19 DLADGKHHTVTRPPPENGYLHIGHAKSICLNFAGIADYQGLCNLRDPTNPPEENILEY 78
 QY 259 VENLKDIETLGIKYDA-VITYSDYFPLMAESLIQGAAYIDDTKQMKRER----- 313
 DB 79 VESIKQVTLGFDWSGEVCYSSDYFPLLYEALIELQGLAYVELTPEQIREYRGTLT 138
 QY 314 MDGISRCNNTVENVLSMKEMVNGTERGQCCVRGLDMQDPNKSRLDPVYVRCNTDP 373
 DB 139 EPQKSPRPDRKSVENNALFEKMAAGEAEGACLRKIDMASPTVWRDVLIVRPAE 198
 QY 374 HHRVSKKYVPTYDFACPFYDALEGVTHALRSSEYHDRAQYRIIDQGLR-RVEIYE 432
 DB 199 HHQGDKKWCICYPMDFTHCISDALEGITHSICLTFQONRLYDMVLNITIPCHPRQYE 258
 QY 433 FSLNMTYLLSKRKLMPVQNKYEDWTPRPPTVQGIYRGKLVKVALIQFLIQGASK 492
 DB 259 FSRNLNLTAVMSKRLQLVTEKLVGTGWDPRMPTISGLRRRGTPSAIRFCRIGVTK 318
 QY 493 NLNLMENDKMTINKIIDIYPCARHTAVLKQDQVY-----FTLTNGPEEPFYRI 541
 DB 319 QENNIEYSALSCIRDDLINENAPRAAVLDPVKLVIENTFAAGYVETLTLANHPKPEH-- 376

QY 542 LPRHKEGAGKATTPANRIWLDYAD-AAAINK-----GEEVTLMDWGNALVKEIKV 593
 DB 377 -----GDREVPFTRFELMIREDFREANKKYKRLVIGKEVRLL--GAYVIKARI 424
 QY 594 ---ESGVIT-----ELVGEHLHESGVTKTKITWLADEELVPLSLVEFDYLISK 642
 DB 425 EKDEQGNITTFCSYDPELTGKNPADG--RKVKGVHWS--AEKGVPAEPFLYERLFTVP 481
 QY 643 KLEBEDEDLDLNPCTREIPALGDANMRNKKRGIIQLERKGYRCAPFIRSKPVVL 702
 DB 482 NPGAADNPAETINPESLVKGVGYPEPSLVEAKPEFGVPERMGYFCADN---KOSSPAL 538

RESULT 14
 SYO_HAEIN STANDARD; PRT; 557 AA.
 ID SYO_HAEIN STANDARD; PRT; 557 AA.
 AC P43831;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
 DE (GlnRS).
 GN GINS OR H11354.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Klebschmann R.D., Adame M.D., White O., Clayton R.A., Kirkness E.F.,
 Kerslavage A.R., Bule C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
 Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Frazer C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd.";
 RL Science 269:496-512(1995).
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
 CC -1- diphosphate + L-glutaminyl-tRNA(Gln).
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL/ U32814; AAC23001.1; -;
 DR PIR/ G64118; G64118.
 DR HSSP/ P00962; 1GTR.
 DR TIGR/ H11354;
 DR HAMAP/ MF_00126; -; 1.
 DR InterPro/ IPR004514; GINS.
 DR InterPro/ IPR000924; Gln-tRNA-synt_1c.
 DR InterPro/ IPR001412; tRNA-synt_1.
 DR Pfam/ PF00749; tRNA-synt_1c; 1.
 DR Pfam/ PF03950; tRNA-synt_1c; 1.
 DR PRINTS/ PRO0967; TRNASYNTGDU.
 DR TIGRFAMs/ TIGR00440; glns; 1.
 DR PROSITE/ PS00178; AA TRNA_LIGASE_I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;
 KW Complete proteome.

FT SITE 42 52 "HIGH" REGION.
 FT SITE 277 281 "KMSKS" REGION.
 FT BINDING 280 280 ATP (BY SIMILARITY).
 SQ SEQUENCE 557 AA; 64078 MW; 4FF1452D57A29F12 CMC64;
 Query Match 19.0%; Score 710.5; DB 1; Length 557;
 Best Local Similarity 34.5%; Pred. No. 8.7e-42;
 Matches 182; Conservative 93; Mismatches 214; Indels 39; Gaps 13;
 QY 199 DLPAKGVKVCVRAPAPSPGSLHTGHAKALLNKYFAPRYOGRILVPPDTPNPKESNEF 258
 DB 27 DLASGKHKSVMTRPPEPNGLYHTGHAKSLICLNGLAKAYGLCNLRDTPNPKVEDVEY 86
 QY 259 VENILKDIETLGIKYDA-VTYSDFPKLMEASLTKOGKAYIDTPKQMKRER----- 313
 DB 87 VDSIKADVEMGTFKMEGEPRIASDYFDALYGYAVELLKIGLAVYDELSPDMREYRGILT 146
 QY 314 MDGIESRCRNNTVEENISLWKEMVNGTERGMOCCVRGLDMQDNKSLRDPVYRCNTDP 373
 DB 147 EPGKNSPYRDRITRENIALFEKMGNGEFAEGKASLRKIDWASPFWVREPVYIRIKFSS 206
 QY 374 HHRVGSKKYVPTYPDFACPFDALEGVTHALRSSEYHNRNQQYRIIDQGLRR--VEIY 431
 DB 207 HHQTDGKWCYIPMYDFTHCISDAIERITHSICTELEFDQNRRLYDVMLENISIERPLPRQY 266
 QY 432 EFSRLNMYVTLISKRKLIMFQNKVDEWTPRPPTVOGIYRGLKVEALIOFLIOGAS 491
 DB 267 EFSRLNLEGITLSGRKLILKYNDEIVDQNDPRMPTISGLRRKGTASLAEFCRIGVT 326
 QY 492 KNLIMEMDKLMTINKKIIDPVCARHTAVLKDQVIFLTNGPEEPFVRLIPRHKEGGA 551
 DB 327 KQDVVEYSALAEACIRDELNENAPRAVAVIDPVRVVI---ENFSEAVLTAPNHNRPDL 383
 QY 552 GKATTPANRIWLDYAD-AAAINK-----GEEVTLMDWGNALVKEIKV---SGVIT 599
 DB 384 GERDLPTKELEYIRADFREBANQYKRLVIGKEVRLL--NAVYIAERVEKXANGKIT 440
 QY 600 -----ELVGEHLHESGVTKTKITWLADEELVPLSLVEFDYLISKLEBEDFL 651
 DB 441 TIFCTYDEPTIGKNPADG--RKVKGVHWSAVNN-HPAERRLYDRLTVPNPGAEIDIE 497
 QY 652 DNLNPTCREIPALGDANMRNKKRGIIQLERKGYRCAPFIRSKPV 699
 DB 498 SVLNPNSLVIRKQGVVEQSLNAAEAKGQFPEREGYFCADS---KDSRP 542

RESULT 15
 SYO_CLOPE STANDARD; PRT; 552 AA.
 ID SYO_CLOPE STANDARD; PRT; 552 AA.
 AC O8XMP3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
 DE (GlnRS).
 GN GINS OR GLTX OR CPB0645.
 OS Clostridium perfringens.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1502;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=13 / Type A;
 RX MEDLINE=21664373; PubMed=11792842;
 RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic
 RT flesh-eater.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
 CC -1- diphosphate + L-glutaminyl-tRNA(Gln).
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

```
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003187; BAB80351.1; -
DR HAMAP; MF_00126; -; 1.
DR InterPro; IPR004514; Glns.
DR InterPro; IPR000924; Glu-tRNA-synt_1c.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00749; tRNA-synt_1c; 1.
DR Pfam; PF03950; tRNA-synt_1c_C; 1.
DR PRINTS; PR00987; TRNASYNTHGU.
DR TIGRfam; TIGR00440; glns; 1.
DR PROSITE; PS00178; AA-tRNA_LIGASE_1; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;
KW Complete proteome.
FT SITE 33 43 "HIGH" REGION.
FT SITE 266 270 "KMSKS" REGION.
FT BINDING 269 269 ATP (BY SIMILARITY).
SQ SEQUENCE 552 AA; 63933 MW; A349EA7211F66FID CRC64;

Query Match 18.9%; Score 705.5; DB 1; Length 552;
Best Local Similarity 33.5%; Pred. No. 1.9e-41;
Matches 177; Conservative 95; Mismatches 209; Indels 47; Gaps 12;

QY 199 DLPGAQVGVKVRPAPEPSGYLHGHAKALINKYFAERYQRLIVRPDDTNPSEKSENF 258
DB 18 DLDSKHDHTIITRFPEEPENGYLHGHAKSIVLNFELGKEFNGRTNLRFDDTNPTEKDETX 77
QY 259 VENLLKDIETIGIKADAVTYSDFPKLMEAMESLIKQKAYIDDTPKQOMKER---M 314
DB 78 VESTIEDVHWLGNSELSHFASNYFDEMVKRALLIKKAKAVCDLPETIEIKYRGTLTE 137
QY 315 DGIESRCNNNTVEENLSLMEKVNNGTERGMOCVAGKLDMDPNKSLRDPVYRCNTDPH 374
DB 138 PGKSPYNNRISIEENLDLFEERRKGEFEDGSKVLAKIDMSSPNINFRDPIIYRIAHASH 197
QY 375 HRVGSKYVVPYTFACPFVDALSEGVTALRSSEYHNRNAQYRILQDMGLRVEI--- 430
DB 198 HNTGDKMCIIYPMYDPAHPLDIALSIGTISICTLEFAD----HRPLYDMFVCECMESVP 252
QY 431 --YESSRLNMYTTLSSKRLMLFVONKKVDEWTDPRFPTVOGIVRGKLVKLEIIOFILO 488
DB 253 ROIEFARLINTNTVNSKKRLKQLVDEGIVDGMDPRMPTVAGLRRGYTPKSIKRFCKAI 312
QY 489 GASKULINEMWDKLTINKKIIDPVCAHTAVLKDQVYFTLTNGPE--EPFVRLIPRHK 547
DB 313 GVAKADSTVDSOMLEHFIREDLQETAPRAMAVINPLKV--ITNYPEGESSEILEIENPK 370
QY 548 FEGAGKKAATTPANRIWLDYADA-----AINKGEVTLMDGNAIVK--EIKVESG 596
DB 371 DESAGKRAVTFESREVIYIREDFMENP PKKYFRLPGNEVRLK--GAYFYKCNIEVIKDENG 428
QY 597 VITELVGLHLEGSVKT-----TKLKITWLADI BELVPLSLVEFDYLISKKLE- 645
DB 429 EVT---EIHCTYDEPETSGETGFGTKVKVGTIHWV--DANNCTPAEFRLYEPLIIDCEEN 483
QY 646 EDEDFLDNINPCTRREIPALGDANNRNIRGEIIOLEKGYRCDA PF 693
DB 484 EGNHFLBQINPNSLTICGFIKPSAKDAKPODKYQLFRHGYNVDPNF 531
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2004, 16:58:19 ; Search time 196 Seconds
(without alignments)
941.165 Million cell updates/sec

Title: US-09-831-683B-10
Perfect score: 3734
Sequence: 1 MEALSPKSDSPISIIICAA.....SSKEVLFALPDGRQASLS 715

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_plant:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacterioph:*
17: sp_archaeoph:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3384	90.6	713	10 Q8S5M6	Q8S5M6 oryza sativ
2	3196	85.6	715	10 Q9L1Z8	Q9L1Z8 oryza sativ
3	2657	71.2	719	10 O82462	O82462 arabidopsis
4	2644.5	70.8	728	10 O65253	O65253 arabidopsis
5	2141	57.3	429	10 O8H8J4	O8H8J4 oryza sativ
6	1529	40.9	1714	5 O35TJ3	O35TJ3 drosophila
7	1516	40.6	1022	11 O8CGC7	O8CGC7 mus musculu
8	1339	35.9	1149	5 Q23J15	Q23J15 caenorhabdi
9	1287.5	34.5	717	4 O8NAJ6	O8NAJ6 homo sapien
10	1276	34.2	863	5 O8IDK7	O8IDK7 plasmodium
11	1139	30.5	642	5 O8SSR4	O8SSR4 encephalito
12	750	20.1	554	16 O8FJW4	O8FJW4 escherichia
13	685.5	18.4	918	5 O8E10	O8E10 plasmodium
14	673.5	18.0	791	10 Q9LQ07	Q9LQ07 oryza sativ
15	669	17.9	556	16 O8EGZ6	O8EGZ6 shewanella
16	653	17.5	549	16 Q8D2R6	Q8D2R6 wigglewort

17	651.5	17.4	580	17 Q8U064	Q8U064 pyrococcus
18	645.5	17.3	795	10 Q8W4F3	Q8W4F3 arabidopsis
19	610	16.3	248	5 O15562	O15562 nosema locu
20	599	16.0	786	10 Q9PR14	Q9PR14 arabidopsis
21	566.5	15.2	606	11 Q8R1V9	Q8R1V9 mus musculu
22	566.5	15.2	775	11 Q8B021	Q8B021 mus musculu
23	566.5	15.2	775	11 Q8BML9	Q8BML9 mus musculu
24	562.5	15.1	612	4 Q96AW5	Q96AW5 homo sapien
25	561.5	15.0	608	4 Q9B3A5	Q9B3A5 homo sapien
26	561.5	15.0	612	4 Q9B0Z3	Q9B0Z3 homo sapien
27	548	14.7	566	17 Q971D0	Q971D0 sulfolobus
28	540.5	14.5	579	17 Q8T7S2	Q8T7S2 methanobarc
29	529.5	14.2	571	17 Q8TXB7	Q8TXB7 methanopyru
30	526	14.1	697	5 Q8SR10	Q8SR10 encephalito
31	510.5	13.7	571	17 Q8PW52	Q8PW52 methanobarc
32	502.5	13.5	570	17 Q8ZU33	Q8ZU33 pyrobaculum
33	492	13.2	250	5 O15563	O15563 trichomonas
34	482.5	12.9	305	5 O15561	O15561 nosema locu
35	464	12.4	548	17 Q979Q0	Q979Q0 thermoplasma
36	438	11.7	531	5 Q95WAL	Q95WAL schistosoma
37	431	11.5	581	1 Q9P9C2	Q9P9C2 uncultured
38	414.5	11.1	394	16 Q8DFN1	Q8DFN1 vibrio vuln
39	381.5	10.2	240	11 Q8BTR1	Q8BTR1 mus musculu
40	381.5	10.2	586	17 Q9HQ11	Q9HQ11 halobacteri
41	319.5	8.6	168	11 Q9CRF9	Q9CRF9 mus musculu
42	288.5	7.7	147	16 Q8BPN0	Q8BPN0 vibrio vuln
43	277.5	7.4	481	16 Q8RB93	Q8RB93 thermoplasma
44	275.5	7.4	485	16 Q97KC9	Q97KC9 clostridium
45	261.5	7.0	489	2 Q8GDY9	Q8GDY9 heliobacilli

ALIGNMENTS

RESULT 1

ID	Q8S5M6	PRELIMINARY;	PRT;	713 AA.
AC	Q8S5M6;			
DT	01-JUN-2002 (TREMBL)	21, Created		
DT	01-JUN-2002 (TREMBL)	21, Last sequence update		
DT	01-MAR-2003 (TREMBL)	23, Last annotation update		
DE	Putative glutamyl-CRNa synthetase.			
GN	OJ1003C07.13.			
OS	Oryza sativa (japonica cultivar-group).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	SpERMatozoa; Magnoliophyta; Liliopsida; Poales; Poaceae;			
OC	Ehharctoidea; Oryzae; Oryza.			
OX	NCBI_TaxID=39947;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRATIN=CV. Nipponbare;			
RA	McCombie W.R., Spiegel L., de la Bastide M., Preston R., Ferraro K.,			
RA	Kuit K., Nascimento L., Zucavert T., Balija V., Bell M., Baker J.,			
RA	Santos L., Miller B., Katzenberger P., Muller S., King L., Yang C.,			
RA	Dike S., O'Shaughnessy A., Palmer L., Dedhia N.;			
RT	"Genomic sequence for Oryza sativa, Nipponbare strain, clone			
RT	OJ1003C07, from chromosome 10, complete sequence."			
RL	Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AC113315; AM08833.1; -			
DR	Genome; Q8S5M6; -			
DR	InterPro; IPR004526; G1X arch.			
DR	InterPro; IPR00924; G1X arch.			
DR	InterPro; IPR00412; CRNA-synt_1c.			
DR	InterPro; IPR00749; CRNA-synt_1c.			
DR	Pfam; PF03950; CRNA-synt_1c; 1.			
DR	TIGRFAMs; TIGR00463; G1X arch; 1.			
DR	PROSITE; PS00178; AA TRNA_LIGASE_1; 1.			
KM	Aminoacyl-tRNA synthetase.			
SC	SEQUENCE 713 AA; 80887 MW; 3E7C2020430DE803 CRC64;			
Query Match	90.6%; Score 3384; DB 10; Length 713;			
Best Local Similarity	89.1%; Pred. No. 2.5e-234;			
Matches 637; Conservative 44; Mismatches 32; Indels 2; Gaps 1;				

QY 1 MEALSFSDSPSPISIIICAAKLVGLPTLNHSLAAGSAPTLQFASGESLHGVPNPIILYIA 60
 Db 1 MEAKLSFSDSPSPISIIISAAKAVAGVSLSDPSLGAAGSAPVLCFSSGSESLRGINPILLEYIA 60
 QY 61 RGASIASISGKNDIEFGHVEMLEVPAPFLTSGSEENCLFVGDGLASTRTFLVGHGLTIA 120
 Db 61 Q-SSPSLHGRDALIESGHVEMLEVPAPFLTSGSEENCLFVGDGLASTRTFLVGHGLTIA 118
 QY 121 DIAVMSNLAGIGQWESLRSKSKYQNLVRWFNSIDSEYKALNEVVAAPVGRKIGKSPA 180
 Db 119 DITVMSNLAGIGQWESLRSKSKYQNLVRWFNSIDSDYKDLNENIYAIVGRKIGKSPA 178
 QY 181 PSLEKXVHDSKDPASPEVDLPAGKVGKVCVRPAPSPSGYLHGHAKAALLNKYFAERYQ 240
 Db 179 PNLKXKVDSDKDPASPEVDLPAGKVGKVCVRPAPSPSGYLHGHAKAALLNKYFAERYQ 238
 QY 241 RLIVRPDDTNPSSKESNEFVENLKDIEFLGIKYDAVVTSDYFPRKLMEMASLIKQKAY 300
 Db 239 RLIVRPDDTNPSSKESNEFVENLKDIEFLGIKYDAVVTSDYFPRKLMEMASLIKQKAY 298
 QY 301 IDDPKEQMRKBERMDGIESRCRNNTVEENLSLMEKENVGTERGMQCCVRGKLDMDPNKS 360
 Db 299 VDDTPKEQMRBERMDGIESRCRNNTVEENLSLMEKENVGTERGMQCCVRGKLDMDPNKS 358
 QY 361 LRDPVYRCNTDPPHHRVGSKYKVPYTPDFACPVDLLEGVTHALRSSEYHNRNAQYRI 420
 Db 359 LRDPVYRCNTDPPHHRVGSKYKVPYTPDFACPVDLLEGVTHALRSSEYHNRNAQYRI 418
 QY 421 QDMGRARVEIYEFSSLNVTLLSKRKLMPQNKVDMDTPRPTVOGIVRBLKXEA 480
 Db 419 QDMGRARVEIYEFSSLNVTLLSKRKLMPQNKVDMDTPRPTVOGIVRBLKXEA 478
 QY 481 LIQFLIOGASKNLNMEMDKLMTINKKIIPVCARHTAVLKDQRYIFLTITNGPEEPVR 540
 Db 479 LIQFLIOGASKNLNMEMDKLMTINKKIIPVCARHTAVLKDQRYIFLTITNGPEEPVR 538
 QY 541 ILPRHKKFEGAGKATTFANRIWLDYADAAAIKGEVTLMDGNAIIVEIKVESGVITE 600
 Db 539 ILPRHKKFEGAGKATTFANRIWLDYADAAAIKGEVTLMDGNAIIVEIKVESGVITE 598
 QY 601 LVGEHLHESGVKTKTKITWLADEIPLSLVEEDYILSKKLEBEDDFDLNLPCTR 660
 Db 599 LVGEHLHESGVKTKTKITWLADEIPLSLVEEDYILSKKLEBEDDFDLNLPCTR 658
 QY 661 EIPALGDANMRNIRKGEIIOLEKGYRCADAPFIRSSKPVVFAIPDGRQOAS 715
 Db 659 EIPALGDANMRNIRKGEIIOLEKGYRCADAPFIRSSKPVVFAIPDGRQOAS 713

RESULT 2
 Q9L128 PRELIMINARY; PRT; 715 AA.

AC Q9L128; 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DE 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 OS Similar to glutamy-tRNA synthetase.
 OC Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OC NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (Ga3) genomic DNA, chromosome 1, PAC
 clone: P063B08."
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP001081; BAA90375.1; -.
 DR HSBP; P00962; IGTB.
 DR Gramene; Q9L128; -.

DR InterPro; IPR004526; GltX arch.
 DR InterPro; IPR000924; GltX ERNA-synt_1c.
 DR InterPro; IPR004046; GST_Cterm.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF00043; GST_C; 1.
 DR Pfam; PF00749; tRNA-synt_1c; 1.
 DR Pfam; PF03950; tRNA-synt_1c; 1.
 DR PRINTS; PR00987; TRNASYNTHGU.
 DR TIGRFAMs; TIGR00463; gltX_arch; 1.
 DR PROSITE; PS00178; AA_TRNA_LIGASE_1; 1.
 KW Aminoacyl-tRNA synthetase.
 SQ SEQUENCE 715 AA; 81053 MW; 9B2E55B019B62D CRC64;

Query Match 85.6%; Score 3196; DB 10; Length 715;
 Best Local Similarity 82.6%; Pred. No. 8; Le-221;
 Matches 590; Conservative 74; Mismatches 48; Indels 2; Gaps 2;

QY 1 MEALSFSDSPSPISIIICAAKLVGLPTLNHSLAAGSAPTLQFASGESLHGVPNPIILYIA 60
 Db 1 MEPLAFPODSPPLIICAAKAVAGVSLTLDKLSGASAPTLHSGSDPIHGCVNITLIRYIA 60
 QY 61 RGASIASISGKNDIEFGHVEMLEVPAPFLTSGSEENCLFVGDGLASTRTFLVGHGLTIA 120
 Db 61 RVASVTSFYGGDALIQAAVYDQMDLYAPVILSGSEFEAAACSLDGYLASRTFLVYGSIA 120
 QY 121 DIAVMSNLAGIGQWESLRSKSKYQNLVRWFNSIDSEYKALNEVVAAPVGRKIGKSPA 180
 Db 121 DIVVMSNLAGIGQWESLRSKSKYQNLVRWFNSI-ADYSDTLDEVSAAYVGRKIGKSPA 179
 QY 181 PSLEKXVHDSK-DPSAPEVDLPAGKVGKVCVRPAPSPSGYLHGHAKAALLNKYFAERYQ 239
 Db 180 PSLEKXVHDSKQNTSIEVDLPAGKAVSVCVRPAPSPSGYLHGHAKAALLNKYFAERYK 239
 QY 240 GRLIVRPDDTNPSSKESNEFVENLKDIEFLGIKYDAVVTSDYFPRKLMEMASLIKQKGA 299
 Db 240 GRLIVRPDDTNPSSKESNEFVENLKDIEFLGIKYDAVVTSDYFPRKLMEMASLIKQKGA 299
 QY 300 YIDDPKEQMRKBERMDGIESRCRNNTVEENLSLMEKENVGTERGMQCCVRGKLDMDPNK 359
 Db 300 YVDDTPKEQMRBERMDGIESRCRNNTVODNLSLMEKENVGTERGMQCCVRGKLDMDPNK 359
 QY 360 SLRDPVYRCNTDPPHHRVGSKYKVPYTPDFACPVDLLEGVTHALRSSEYHNRNAQYRI 419
 Db 360 SLRDPVYRCNTDPPHHRVGSKYKVPYTPDFACPVDLLEGVTHALRSSEYHNRNAQYRI 419
 QY 420 LQDMGLARVEIYEFSSLNVTLLSKRKLMPQNKVVEDTDRPPTVOGIVRBLKXEA 479
 Db 420 LQDMGLARVEIYEFSSLNVTLLSKRKLMPQNKVVEDTDRPPTVOGIVRBLKXEA 479
 QY 480 ALIOPFLIOGASKNLNMEMDKLMTINKKIIPVCARHTAVLKDQRYIFLTITNGPEEPVR 539
 Db 480 ALVQFLIOGASKNLNMEMDKLMTINKKIIPVCGHHTAVLKDQRYIFLTITNGPEEPVR 539
 QY 540 RILPRHKKFEGAGKATTFANRIWLDYADAAAIKGEVTLMDGNAIIVEIKVESGVITE 599
 Db 540 RILPRHKKFEGAGKATTFANRIWLDYADAAAIKGEVTLMDGNAIIVEIKVESGVITE 599
 QY 600 ELVGEHLHESGVKTKTKITWLADEIPLSLVEEDYILSKKLEBEDDFDLNLPCTR 659
 Db 600 ELVGEHLHESGVKTKTKITWLADEIPLSLVEEDYILSKKLEBEDDFDLNLPCTR 659
 QY 660 REIPALGDANMRNIRKGEIIOLEKGYRCADAPFIRSSKPVVFAIPDGRQOAS 713
 Db 660 QEPALGDANMRNIRKGEIIOLEKGYRCADAPFIRSSKPVVFAIPDGRQOAS 713

RESULT 3
 O82462 PRELIMINARY; PRT; 719 AA.

AC O82462; 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Glutamy1-tRNA synthetase.
 GN AT5G26707.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98438737; PubMed=9765600;
 RA Day I.S., Golovkin M., Reddy A.S.;
 RT "Cloning of the cDNA for glutamy1-tRNA synthetase from Arabidopsis
 thaliana.";
 RL Biochim. Biophys. Acta 1399:219-224 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shimizu P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shimizu P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF067773; AAC36469.1; -;
 DR EMBL; AY095952; AAC20443.1; -;
 DR EMBL; BT000248; AAN15567.1; -;
 DR HSSP; P00962; 1GTR.
 DR InterPro; IPR004526; GLTX arch.
 DR InterPro; IPR000924; GLU tRNA-synt_1c.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF00749; tRNA-synt_1c; 1.
 DR Pfam; PF03950; tRNA-synt_1c; 1.
 DR PRINTS; PRO0987; TRNASYNTGTLU.
 DR TIGRFAMs; TIGR00463; gltx arch; 1.
 DR PROSITE; PS00178; AA tRNA_LIGASE_I; 1.
 KW Aminoacyl-tRNA synthetase.
 SQ SEQUENCE 719 AA; 81064 MW; 0778C243219DA24C CRC64;
 Query Match 71.2%; Score 2657; DB 10; Length 719;
 Best Local Similarity 69.0%; Pred. No. 4,2e-182;
 Matches 491; Conservative 106; Mismatches 107; Indels 8; Gaps 5;

QY 5 LSFKSDSPISITICAKVLPLTINSLAAGAPTLOFAPAGESIHGVPITILYARGA- 63
 DB 6 LSPFSPESPLSVIVLSTLSASPVTIDSSAATAVTPSFVFSGDKRKINGATVLLRYGRSAK 65
 QY 64 SIASISGKNDLEFGHVWLEAPYFPLSGSEFENACLFVDGFLASRTLVNGHGLTIAID 123
 DB 66 KLDPFYGNNAFDSQIDEMWVYASVFSGSEFENACGVNDYKLESTYLVHSHSIAIVA 125
 QY 124 VWSNLAGIGQWESLRSKSKYQNTLVWPNFNSIDSEYKALNEVVAFVGRGIGKS-PAPS 182
 DB 126 IWSALAGTQGWESLRSKSKYQNTLVWPNFNSIDSEYKALNEVVAFVGRGIGKS-PAPS 184
 QY 183 LKEKXV---DSKPSADEVDLPGAKGVKCVRFAPFSPGYLHGHAKAALLINKYFARY 238
 DB 185 SKDSQAAVKGQDGKGPEDVLPFAELIGKVKLRFAPFSPGYLHGHAKAALLINKYFARY 244
 QY 239 OGRLIVRPDTPNPSKESNEFENLLKQETLGIKYDATYTSVDPKLMEMAEELIKGK 298
 DB 245 QGEVIVRDDTNPAPESNEFDNLVKDITGIGIKETKYTSYDFPELMDAEELIKGK 304
 QY 299 AYIDTPEQMRKERMDIESRCRNATVEENLSTLKEWVNGTERMQCCVAGKLDMDOPN 358

DB 305 AYVDTPREQOMKERMDIGSKCRNHSVEENLKLWEMIAIGSERGLQCCVAGKFNMDOPN 364
 QY KSLRDPVYRCNTDPHHRVSGKYKVPYDPACFPVDALEGVTHALRSSEYHNRNAYYR 418
 DB 365 KAMDVPYRCNPNPMSHRIIGKIKIPTYDACPVSLSBGITHALRSSEYHNRNAYYR 424
 QY 419 ILDDWGLRRVEIYFESRLNMYTLLSKRKLMPYQNKVDEMTDPRPPTVQGIYARGIKV 478
 DB 425 VLEMGARQVQVLYFESRLNMYTLLSKRKLMPYQNKVDEMTDPRPPTVQGIYARGIKI 484
 QY 479 EALLIQTIIQCGASQNLNMEMDKIMTINKIIPVCAHRYVLKDQRYIFTLNGPEEPF 538
 DB 485 EALLIQTIIQCGASQNLNMEMDKIMTINKIIPVCAHRYVLKDQRYIFTLNGPEEPF 544
 QY 539 VRIIPRHKPEGAGKKAATTFANRIMLVADAAANKGEVTLMDGNATVVEI-KVESGV 597
 DB 545 VRIIPRHKPEGAGKKAATTFANRIMLVADAAANKGEVTLMDGNATVVEI-KVESGV 604
 QY 598 ITTELVEGLHLEGSVYKTKLKITWMLADIEELVPLSLVEFDYLLISKKLEDEDFLDNLNPC 657
 DB 605 VTALSGVNLNQGASVYKTKLKITWMLADIEELVPLSLVEFDYLLISKKLEDEDFLDNLN 664
 QY 658 TRRIIPALGDANMENIRGEIITQLEKGYRCDAFTRSSKPVVLPALPDGR 709
 DB 665 TKKETLALGDSNMRLKCGDVITQLEKGYRCDAFTRSSKPVVLPALPDGR 716

RESULT 4
 ID 065253 PRELIMINARY; PRT; 728 AA.
 AC 065253;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, last annotation update)
 DB F21E10.12 protein.
 GN F21E10.12.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Davidson S., Rohlfing T., David M., O'Brian D.;
 RC STRAIN=cv. Columbia;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Washu;
 RC STRAIN=cv. Columbia;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Waterson R.;
 RC STRAIN=cv. Columbia;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Wilson R.;
 RC STRAIN=cv. Columbia;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF058914; AAC13597.1; -;
 DR HSSP; P00962; 1GTR.
 DR InterPro; IPR004526; GLTX arch.
 DR InterPro; IPR000924; GLU tRNA-synt_1c.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF00749; tRNA-synt_1c; 1.
 DR Pfam; PF03950; tRNA-synt_1c; 1.
 DR PRINTS; PRO0987; TRNASYNTGTLU.
 DR TIGRFAMs; TIGR00463; gltx arch; 1.
 DR PROSITE; PS00178; AA tRNA_LIGASE_I; 1.

SQ SEQUENCE 728 AA; 82066 MM; F7D42BAEAD5A5E93 CRC64;
 Query Match 70.8%; Score 2644.5; DB 10; Length 728;
 Best Local Similarity 68.2%; Pred. No. 3.4e-181;
 Matches 492; Conservative 105; Mismatches 107; Indels 17; Gaps 6;
 5 LSPFSDSPPIITICAAKVGPLPTINHLAASAPLTOPAGSESLHGNPIITLYARA- 63
 6 LSPFSPSPPLSVIALSLASPVITIDSSAAATVPSFVSOGKRLNGATVLLRYGRSAK 65
 64 SIASISGNDIEFGHV-----EWLEAYAPTLGSEFENACLTFDGFASRTPVAG 114
 66 KLDPFGNNAFDSQSVSLICINMKIDEVNDVASVSSSEFENACGRDKLSESTFLVG 125
 115 HGLTIADIVASNLAGIQRWESLRSKKYONLVWMSIDSEYKALNEVVAAPVGRG 174
 126 HSLSIADVAIVSALAGTQWRWSLRSKKYQSLVWFSIDSEYEVNLKVLATVYV-KKG 184
 175 IGKS-PASLSKEVH-----DSKDPSPAEVYDLPGAIVGKVCVRPAPEEGYLHIGAKAL 229
 185 SGKPVAAERKSDSQAVKGDQDKGKPEVDLPBAEIGKVKLRFAPBEPGYLHIGAKAL 244
 230 LNKYFAERYOGLIYRPDTPNPSKSENFVENLKDIEITGIXYAVYTSDPYPKLMM 289
 245 LNKTFAEHYQGEVIVRFDPDTPNPAKSNSEFVDNLVKDITGLIKYKAVYTSDPYFELMDM 304
 290 AESLIKQKAYIDTTPKEOMRKERMKGIESRCNNVTEENLSLMEKNVNGTERGMQCVR 349
 305 AEKLMREKAYVDTPREOMOKERMDGIDSKCRNHSVEENLKLMEWLAGSERGLQCCVR 364
 350 GKLMDQDNKSLRDPVYRRCNTDPHHRVGSCKKYPTYPDFACPFDALGEGYTHALRSSEY 409
 365 GKFNQODPNKMRDVPVYRCNPMSSHRIQDKYKIPYDFACPFVDSLEGYTHALRSSEY 424
 410 HDRNAOYRIIDQMGRLRVEIYEFSLNMVYLLSKRLMVFONKRYEDMTDPFPVQ 469
 425 HDRNAQYRVKVEDMGLROYLTFESRLNLVFLLSKRLMVFQGLVDGMDPFPVQ 484
 470 GIVRGLKVEALIQFILOQASKNLNMWMDKLMTINKKIIDPVCARHTAVLKQORVLT 529
 485 GIVRGLKIEALIQFILOQASKNLNMWMDKLMTINKKIIDPVCARHTAVLAERVAFT 544
 530 LTNGEPEPVRILPPHKKFEGAGKATTPANRIWLDVADAANKGEEVYTLMDGNAIVK 589
 545 LTGDPDEPFRMIPGHKKFEGAGKATTPFKSIWLEADASAI SVGEETLMDGNAIVK 604
 590 EI-KYESGIVTELVEGLHLEGSVKTTKLKITWLADIEELVPLSLVEFPYLSKKLSEDE 648
 605 EITKDEBGRVTLASVNLQGSVKTKTKLKITWLPTNELVNLITFEDLITKKLEDDO 664
 649 DFLDNLNCTRRREIPALGDANMRNIRKGEIIOLEKRGYRCDAPIRSSKPVLPFAIPDG 708
 665 EVADVNTNKTETALDGSNMNRNKKCGVIOLEKRGYRCVPIVSKSKPIVLPSPIDG 724
 QY 709 R 709
 DB 725 R 725
 RESULT 5
 Q8H8J4 PRELIMINARY; PRT; 429 AA.
 AC Q8H8J4;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Putative RNA synthetase.
 GN OSJBA0083M08.1.
 OS Oryza sativa (japonica cultivar-group).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC Euphorbiaceae; Oryzae; Oryza.
 NCBI_TaxID=39947;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Buehl C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
 RA Overton II L.L., Telstrin T., Kim M.M., Bera J.J., Jin S.S.,
 RA Fadron D.W., Tallon L.J., Koo H., Zismann V., Hejiao J., Blunt S.,
 RA Vanaken S.S., Riedmiller S.B., Uteback T.T., Feldblum T.V.,
 RA Yang Q.Q., Haas B.U., Suh B.B., Peterson J.J., Quackenbush J.,
 RA White O., Salzberg S.L., Frazer C.M.;
 RT "Oryza sativa chromosome 10 BAC OSJBA0083M08 genomic sequence."
 RW Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 KR EMBL: AC096691; AAN08648.1; -
 KW Aminocycl-tRNA synthetase.
 SQ SEQUENCE 429 AA; 49743 MM; E1A8B2239486622 CRC64;
 Query Match 57.3%; Score 2141; DB 10; Length 429;
 Best Local Similarity 92.1%; Pred. No. 2.2e-145;
 Matches 395; Conservative 25; Mismatches 9; Indels 0; Gaps 0;
 287 MEMAESLIKQKAYIDTTPKEOMRKERMKGIESRCNNVTEENLSLMEKNVNGTERGMQC 346
 1 MEMAEKLIKQKAYVDTPKEOMRKERMKGIESRCNNVTEENLSLMEKNVNGTERGMQC 60
 347 CVRGLMDQDNKSLRDPVYRRCNTDPHHRVGSCKKYPTYPDFACPFDALGEGYTHALRS 406
 61 CVRGLMDQDNKSLRDPVYRRCNTDPHHRVGSCKKYPTYPDFACPFDALGEGYTHALRS 120
 407 SEYHDRNAOYRIIDQMGRLRVEIYEFSLNMVYLLSKRLMVFONKRYEDMTDPFP 466
 121 SEYHDRNAOYRIIDQMGRLRVEIYEFSLNMVYLLSKRLMVFONKRYEDMTDPFP 180
 467 TVQGIIVRGLKVEALIQFILOQASKNLNMWMDKLMTINKKIIDPVCARHTAVLKQORV 526
 181 TVQGIIVRGLKVEALIQFILOQASKNLNMWMDKLMTINKKIIDPVCARHTAVLKQORV 240
 527 IFTLTNGEPEPVRILPPHKKFEGAGKATTPANRIWLDVADAANKGEEVYTLMDGNA 586
 241 IFTLTNGEPEPVRILPPHKKFEGAGKATTPANRIWLDVADAANKGEEVYTLMDGNA 300
 587 IVEIKYESGIVTELVEGLHLEGSVKTTKLKITWLADIEELVPLSLVEFPYLSKKLSE 646
 301 IVEIKYESGIVTELVEGLHLEGSVKTTKLKITWLADIEELVPLSLVEFPYLSKKLSE 360
 647 DEDFLDNLNCTRRREIPALGDANMRNIRKGEIIOLEKRGYRCDAPIRSSKPVLPFAIP 706
 361 DEDFLDNLNCTRRREIPALGDANMRNIRKGEIIOLEKRGYRCDAPIRSSKPVLPFAIP 420
 QY 707 DGRQOASLS 715
 DB 421 DGRQOASLS 429
 RESULT 6
 Q95TL3 PRELIMINARY; PRT; 1714 AA.
 AC Q95TL3;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Lda2739p
 GN AATS-GLUPRO OR CG5394.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Flise E., George R.,
 RA Gonzalez M., Guatin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,

RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
 RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AY058703; AAL13932.1; -
 DR PLAYbase; FBgn0005674; Aats-glupro.
 DR InterPro; IPR001589; Actbind_actnin.
 DR InterPro; IPR004526; GltX_arch.
 DR InterPro; IPR000924; GltX_arch.
 DR InterPro; IPR004046; GST_Cterm.
 DR InterPro; IPR004154; HGT anticonodon.
 DR InterPro; IPR004499; Pros_fam_1.
 DR InterPro; IPR002314; tRNA-synt_2b.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR InterPro; IPR002316; tRNA-synt_1.
 DR InterPro; IPR006195; tRNA_Ligase_II.
 DR InterPro; IPR000738; WHEP-TRS.
 DR Pfam; PF00043; GST_C_1.
 DR Pfam; PF03123; HGT_anticonodon_1.
 DR Pfam; PF00749; tRNA-synt_1c_1.
 DR Pfam; PF03950; tRNA-synt_1c_C_1.
 DR Pfam; PF00587; tRNA-synt_2b_1.
 DR Pfam; PF00458; WHEP-TRS_6.
 DR PRINTS; PR00987; TRNASYNTHU.
 DR PRINTS; PR01046; TRNASYNTHU.
 DR TIGRFAMs; TIGR00463; gltX_arch; 1.
 DR TIGRFAMs; TIGR00408; pros_fam_1; 1.
 DR PROSITE; PS00178; AA tRNA_Ligase_II; 1.
 DR PROSITE; PS00862; AA tRNA_Ligase_II; 1.
 DR PROSITE; PS00019; ACTININ_1; 1.
 DR PROSITE; PS00762; WHEP-TRS; 6.
 DR SEQUENCE 1714 AA; 189411 MW; 3F8CF3DB128765A8 CRC64;

Query March 40.9%; Score 1529; DB 5; Length 1714;
 Best Local Similarity 44.6%; Pred. No. 1.7e-100; Indels 32; Gaps 13;
 Matches 324; Conservative 129; Mismatches 241;

QY 1 MEALSPSKDPPISIIICAKIVG--LPITINHSLAASAPTLQFASGESL-HGVNPII 56
 DB 1 MSITLKNANNPISGLATAHLINGTVPVEIWMSEERS---LQPDRLVLCSSNNVYL 57
 QY 57 LYIRGASIASLSCNDIEFGHVEMLEAPTLPSGSEFENACLFVDFGLASRTFLVGHG 116
 DB 58 RALRAAPDYKLYGETAERTQIDHMLSFSLT--CEDDISWALSLDLSIAPIVLYLVANK 115
 QY 117 LTIDIAVMSNLAGIGQWESLRSKSKYQNLVRWNSIDSEYKALNVAAPFKRIG 176
 DB 116 LTIDPFLAFENM--HSRYEFLAAKGIPOHVQWYDLTAAQ--PIIQVLSLPEDAKVK 170
 QY 177 KSPAPSLKE--KYHDSKDPAPAEVDLPKAGKVCVAFAPESGYLHIGAKAALLNKY 233
 DB 171 RSPSSKEQTPAKTGERKQ--EGKPYDLPGLAMGKVVAFPPBASGYLHIGAKAALLNOY 229
 QY 234 FAERYQGLIVRPDDTNPSEKSENEFENLKDILETIGIKYDAVYTSDFPKLMEASL 293
 DB 230 YALFQGLTLMRPDDTNPKEKEVEENITLGLDLOLKPDPVFTHTSNYFDLMLDYCVRL 289
 QY 294 IKQKAYIIDDTPKQOMKERBDGISRCNNNTVENLSLMEKMNVTGRGQCCRGKLD 353
 DB 290 IKESKAYVDTPPQOMKLERQORVASANRSNSVEKNLSLMEEMVXGSEKQKVCRAKID 349
 QY 354 MODNKSRLDVPYVRCNTDPHHRVSGSKYKVPYVDPAFPVDALEGVTHALRSSEYHNRN 413
 DB 350 MSSNGCKRDPITLRCKNEPHTGTGKVKYFTPDPAFPIVDALINHTLRTTEHYHRD 409
 QY 414 AQQYRILODMGLRVEIYFSLNMYTLLSKRKLWFOQNKVVDWTPRFPYQGIVR 473
 DB 410 DQFWFIDALTKLRKPYIWSYSLNMTNTVLSSKRLTWFDVSGLVGMDPDPFPYRGIR 469
 QY 474 RGLVEALIQITLLOGASKNLNMEDKLTINKKIIDPVCAKHTAVLKQORVIFTLNG 533
 DB 470 RGMTEGKEKPIIAGGSSKSVFPMWMDIKWAFNKKVIDPIAPRYALKEKRVLYNVGA 529
 QY 534 PEEFVRI-LPRHKKFEGAGKATTFARIMLDVADAANKGEVETLMDGNMIVKRI- 591

DB 530 KVE---RIQVSHPKDESIGKTVLLGPRIYIDVDAEALKEGENATFIMGNILIRKVN 586
 QY 592 KVESGVITTELVEGLHLEGSVXTTKITWLA---DIEELVPLSLVEFPLYSKXKLEDE 648
 DB 587 KQASGNITSDAALNLENKPKFKTLKLTWAVEDDPSAYPPTFCYFDNITISKAVLGDE 646
 QY 649 DFLDNLPCTRRRELIPALGDANMRENIKRGEIIQLERKGYRCADAFIRSS-----KPVVL 702
 DB 647 DFKQFIHKHTRDEVPMIGDPELKKCKKGDIIQLGRGPFKVDVAVAPPSGYTVNPSPIVL 706
 QY 703 FALPDG 708
 DB 707 FSLPDG 712

RESULT 7

ID 08CGC7 PRELIMINARY; PRT; 1022 AA.
 AC 08CGC7;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Similar to glutamyl-prolyl-tRNA synthetase (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N;
 RA Strausberg R.;
 RL Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.
 KW Aminoacyl-tRNA synthetase.
 FT NON TER 1022
 SQ SEQUENCE 1022 AA; 114132 MW; B2EBD862D891AB61 CRC64;

Query March 40.6%; Score 1516; DB 11; Length 1022;
 Best Local Similarity 42.9%; Pred. No. 6.6e-100; Indels 40; Gaps 10;
 Matches 310; Conservative 147; Mismatches 225;

QY 5 LSPSKDPPISIIICAKIVGLPLTINHSLAASAPTLQFASGESLHGVNPIIYARGAS 64
 DB 6 LTVAAGNPPLLEALLAVEHVKGDSI--SVESGKNLRLVSTVAFDTNLSRLRIARIAT 63
 QY 65 IASLSCNDIEFGHVEMLEAPTLPSGSEFENACLFVDFGLASRTFLVGHGLTIADIA 123
 DB 64 TSGLYGTNLMENHTEIDHMLFSATKLSSCDGLTSAINELNHCLSLRTLVGNSLTLDLC 123
 QY 124 VMSNLAGIGQWESLRSKSKYQNLVRWNSIDSEYKALNVAAPFKRIGSKSPASL 183
 DB 124 VMAITLKSAAMQEHKLKONKTLVHVAKWFGFLAEO-----QAF--RSVG-----T 165
 QY 184 KEKHSKDPSPAP-----VDLPKAGKVCVAFAPESGYLHIGAKAALLNKYFA 235
 DB 166 KMDVSGSRATYAPDKQDVGKVELPGLAKMKTAVRPPBASGYLHIGAKAALLNGHYQ 225
 QY 236 ERYQGLIIVRPDDTNPSEKSENEFENLKDILETIGIKYDAVYTSDFPKLMEASL 295
 DB 226 VNFPGKLTIMRPDDTNPKEKEDEPKVILLEDVAMHIIKPDQTYTSDHETIMKYEKLIQ 285
 QY 296 QGKAYIIDDTPKQOMKERMDGISRCNNNTVENLSLMEKMNVTGRGQCCRGKLD 355
 DB 286 EGKAYVDTPPQOMKLERQORVESANRSNSVEKNLSLMEEMVXGSEKQKVCRAKID 345
 QY 356 DPNKSLDVPYVRCNTDPHHRVSGSKYKVPYVDPAFPVDALEGVTHALRSSEYHNRN 415
 DB 346 SNGCKRDPITLRCKIOPHRTGKKNVYFTPDPAFPIVDSIBGYTHALRTTEHYHRDQ 405
 QY 416 YRILQDMGLRVEIYFSLNMYTLLSKRKLWFOQNKVVDWTPRFPYQGIVRG 475
 DB 406 FYWIIEMALGRKPYIWEYSRLNMTNTVLSSKRLTWFDVSGLVGMDPDPFPYRGVLRG 465

QY 476 LKVEALIOPILOOGASKNLNMEDKMTINKKIIDPCARHTAVLKQRYIFLITLNPE 535
DB 466 MTEBELKQPIAQQSSRVNMEDKIWAFNKVLDPAAPRVALLKEVFPVNLDAQE 525
QY 536 EPPVILPRHKKFEGAGKATTFANRIMLDYADANAANKSEEVITMDGNAIKVEI-IVE 594
DB 526 E--MKEVARHRKNPVGKLPWYSPKVPFIEGADATPSEGEVNTINNGNINITIKHNA 583
QY 595 SGVITELVGLHLEGSVKTTKITWLDIIEVLPLS--LVEFDYLISKKULEDEDFLD 652
DB 584 DGKITSLDKANTLENKDYKTKITWLAESTHALPFAVCTYEHLLTKPVLGKDEDFQ 643
QY 653 NLNPTREIRIPALGDAMNRNIRKGEIIOLEKGYRCAPFIRSS-----KPVLPFAIP 706
DB 644 YINKDSKHEMLMDGPCKLKKDGIQLQRRGFICDPQPEYSPYSREAPCILYIP 703
QY 707 DG 708
DB 704 DG 705

RESULT 8
Q23315 PRELIMINARY; PRT; 1149 AA.
AC Q23315;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE ZC434.5 protein.
GN ZC434.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilkinson J.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology."
RT Science 282:2012-2018(1998).
DR EMBL; Z5714; CAB00060.1; -.
DR HSSP; P00962; IGTR.
DR WormPep; ZC434.5; CE06580.
DR InterPro; IPR001589; Acctbind_actnin.
DR InterPro; IPR004526; GLTX_arch.
DR InterPro; IPR000924; GLU_rRNA-synt_1c.
DR InterPro; IPR001412; rRNA-synt_1.
DR InterPro; IPR000738; MHEP-TRS.
DR Pfam; PF00749; rRNA-synt_1c; 1.
DR Pfam; PF03950; rRNA-synt_1c; 1.
DR Pfam; PF00458; MHEP-TRS; 6.
DR PRINTS; PR00987; TRNASYNTGILU.
DR TIGFAMS; TIGR00463; GLTX_arch; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
DR PROSITE; PS00019; ACTININ_1; 1.
SQ SEQUENCE 1149 AA; 125197 MW; CC22744C05C0M433 CRC64;

Query Match 35.9%; Score 1339; DB 5; Length 1149;
Best Local Similarity 39.5%; Pred. No. 4,1e-87;
Matches 292; Conservative 132; Mismatches 243; Indels 72; Gaps 13;

QY 2 EAASFSKDSPPISITCAKVLGLPLTNHSLAASAPTLQFA-----SGESLHGV 52
DB 6 ELVLKANBEOPPYASILALASGSFL-----EKSVQSEKQQLALNLDGELLND 55
QY 53 NPILITVIRGASIA-SLSGKNIDIEGHVWEMLEFAP-----TFLSGSEFENACLVD 103
DB 56 VEIARIIGQSTDAADSLGSSIIDPAVVDLTNFIADAVTKNDYSLLGKDP----- 108

QY 104 GFLASRTFLVGHGTLTIADIAVMSYLAGIQRMESLRKSKYQNIIVRMFNSIDSEYKALN 163
DB 109 -----TKILDNSLTJVDFAIFP-----VAHNNPOLK-----AKPSGIIIDKVLKEPTL 150
QY 164 EVAAPFGKQIGKSP---APSLKEVHDSKDPAPEDVDPGAVGVCFAPFAPSGYL 220
DB 151 AAANFVGLYKSAAPATATATGKKEK-----KDGKVELPFGAKGVVAFPEPAGYL 206
QY 221 HIGAKAALLNKRYFAERYQGLIVRFDPTNPKESKSENFVENLKDITLGIKIDAVTYTS 280
DB 207 HIGAKAALLNQYVQAPEGQILIRFDDTNPAKENAFHEVHKEDLSMLNIVPRAWTHSS 266
QY 281 DYPRKLMEASLTKQKAYIDTTPKQMKRERDGIJSKRNNVTENLSMKEMNGT 340
DB 267 DHFEMLLTMCKLKEGAFVDDTDTETMERERORDSRRSNTPEKNQOLMEEMKGS 326
QY 341 ERGMQCCVRGKLDMDQPNKSLRDPVYRCNTDPHHRGVSKYKYVPTYDFACPFDALGCV 400
DB 327 PKGLTCCVRMKIDMKSNNGARDPITYCKPEEHVIRGLKVPYPTDFTCPYDSEGV 386
QY 401 THAIRSEYHDNRNAQYRYILODMGLRREIYEPFSLNMVYTLISKRLMFWQNKYEDW 460
DB 387 THAIRTEYHNRDQYFFICDGLGRPHIMEVARIAMTNTVMKRLTWFDVGHEGW 446
QY 461 TDPRFPYQGIVRGKLYEALIOPILOOGASKNLNMEDKMTINKKIIDPCARHTAV 520
DB 447 DDPRLPVIRGVWRGLTYVEGLKQPIVAQGSRSVMMEMDKIWFNKKVLDPAAPRYAL 506
QY 521 LKQGRVI-FITLTPNEPFPVILPRHKKFEGAGKATTFANRIMLDYADANAANKSEEV 579
DB 507 DSTPLVSIETSDISDTSNV-SLHPNMAISGDVHKKKLLBQVDNALKEGELVT 565
QY 580 LMDGNAIKVEIKIYKESGVITELVGLHLEGSVKTTKITWLDIE-----ELVPLSIVEF 635
DB 566 FVMGNKIKIGKIEKGAIVTISATLODNTDYKTKTKVTLGVGVKAEAGTIVVTADY 625
QY 636 DYLLISKKULEDEDFLDNLNPTREIRIPALGDAMNRNIRKGEIIOLEKGYRCAPFIR 695
DB 626 DHIISAKIIGDEDEWQKPFINDSVHYTKMGVGEPAIKVKKDIIQIQKGFYIDQPYNP 685
QY 696 SSK-----PVLPFAIPDG 708
DB 686 KSELGVETPULLAIIPDG 704

RESULT 9
Q8NAJ6 PRELIMINARY; PRT; 717 AA.
AC Q8NAJ6;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ35251.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Ishibashi T.; Kanehori K.; Yosida M.; Watanabe S.; Ishida S.; Ono Y.;
RA Hociuta T.; Hiraoka S.; Murakawa K.; Takiguchi S.; Kusano J.;
RA Watanabe M.; Fujimori K.; Tanai H.; Ishida M.; Yamauchi H.; Chiba Y.;
RA Sugiyama T.; Irie R.; Otsuki T.; Sato H.; Wakamatsu A.; Ishii S.;
RA Yamamoto J.; Isono Y.; Kawai-Hio Y.; Saito K.; Nishikawa T.;
RA Kimura K.; Matsuo K.; Nakamura Y.; Sekine M.; Kikuchi H.; Kanda K.;
RA Wagatsuma M.; Takahashi-Fujii A.; Oshima A.; Sugiyama A.; Kawakami B.;
RA Suzuki Y.; Sugano S.; Nagahari K.; Masuno Y.; Nagai K.; Isogai T.;
RT "NEO human cDNA sequencing project."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK092570; BAC03916.1; -.
DR InterPro; IPR004526; GLTX_arch.

DR InterPro; IPR000924; Gln_tRNA-synt_1c.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00749; tRNA-synt_1c; 1.
DR Pfam; PF03950; tRNA-synt_1c; 1.
DR PRINTS; PR00987; TRNASYNTHGU.
DR TIGRFAMs; TIGR00463; glnx_arch; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Hypothetical protein.
SQ SEQUENCE 717 AA; 81797 MW; D78B40974C8D4F79 CRC64;

Query Match 34.5%; Score 1287.5; DB 4; Length 717;
Best Local Similarity 41.2%; Pred. No. 9.9e-84;
Matches 280; Conservative 127; Mismatches 244; Indels 29; Gaps 13;

QY 46 GE-SLHGNNPIILYIARASISLSKNDIEGHVEMLEVPYPTLSGSEFENACLPVDG 104
DB 43 GTTIGSDAICDYITSKCADQSLGGDDAE--TVKEMHIASTTTKENVLEQ-TEKLD 99
QY 105 FLASRTPLVGHGLTIADIAVMSNLGIGQWESLRKSKYQNLVWFNSIDSEYKEALNE 164
DB 100 YIATPTIYVAVNTIADISMAAI-----KLEVEIPAKCTFASRWYITISA--IPSVQ 152
QY 165 VVAALFVGRGIGKSPAPSLKEKVH-----DSKDPAPVDLPGARVGVCFAPPEP 216
DB 153 TTGRVGSLSRMAKQASAPKKEIATKKGEISLDGPESWPE--LPFAEMGVVTRFPPEA 210
QY 217 SGYLIHGAKAALINKYFAERYQGLIYRFDPTNPSKESNEFEVULKDIELTGIKYAV 276
DB 211 SGYMHIGHKAMLYYAKRYQGLIRFDPTNPSKEKEFEPTIIIEIDLAKIGKADLF 270
QY 277 TYSIDYFPELMEMASLIKQKAYIIDPTPEQMRKERMDGISRCRNNTVEENLSLAKEM 336
DB 271 SHTSYFVLIIDYARQMIREGIAPMDNTDQETMKERKERKSKLRNTSPENLIPFL 330
QY 337 VNGTERGQCCVRGLDMQDPNKSIRDPIYVRCNTDPHHRVGSXYKVYPTDYFACPYDA 396
DB 331 CRGEPEVDYCLRAKIDMKSDNGTLRDPVLVRYFPLTHRTGDKYKAVPCVDLACPIDS 390
QY 397 LEGVTHARSSHYHRNAQYRIIQQMGIRKVEIYFSLNMYTLLSKRLMYVQNK 456
DB 391 IEGVTHARTTEYKQDRDOYMWIQQALRIKRYHVLVEFALNEQYTLMSRKLTWVNEHE 450
QY 457 VEDWTPRPPTVOGIVRGLKVEALIOFLQOGASKNLNMEMDGLMTINKKIIDPVCA 516
DB 451 VQGDDEPRPPTVQGVLRGKVEALIOFLQOGASKNLNMEMDGLMTINKKIIDPVCA 510
QY 517 HTAVLKDORVIFTLTNGEPEEPVRLIPRHKKFEGAGKATTPANRIWLDYADAALNKG 576
DB 511 FMGVKWSFVTLTLTN-VQPGFIK-PNHPKDESKGVSDIAGPQVLMLERTDAQLKGE 568
QY 577 EYTLMDGNAIYKEIKVES-GVITELVGBLHESGVKTKTKITWLADIIEVLPLSVEF 635
DB 569 EFTLMRWGSAKVTALRADAGAVGTGEFVNPNGDFKNT-AKVNVVAACPEHNEVVEY 627
QY 636 DYLLSKKLEED--EDFLDNLPCTRRRIIPALGDANNRNIKGGIIQLERGVRCAP 692
DB 638 DDLTKTKLLEEDNFELTSDKHPKMEYTVIVNACKGLKENQIVOLERRGFYRCVP 687
QY 693 FIRSKPVVLPALPDGRQQA 712
DB 688 DTENS-PAKTLIIPDGKKA 706

RESULT 10

Q8IDK7 PRELIMINARY; PRT; 863 AA.
AC Q8IDK7;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE Glutamate--cRNA ligase (EC 6.1.1.17).
GN PF13_0257.

OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN (1)
RP SEQUENCE FROM N.A.
RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,
RA Berriman M., Pain A., Hall N., Actin R., Chillingworth C., Doggett J.,
RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrett B.,
RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AL844509; CAD52614.1; --
KW Ligase.
SQ SEQUENCE 863 AA; 101417 MW; AEC2BD5FA7A292B8 CRC64;

Query Match 39.6%; Score 1276; DB 5; Length 863;
Best Local Similarity 39.6%; Pred. No. 8.8e-83;
Matches 273; Conservative 128; Mismatches 222; Indels 60; Gaps 10;

QY 67 SLGKNDIEGHV-----EWLEVAPFLSGSEFENACLPVDGFLASRTPLVGHGL 117
DB 144 NISEKIDPKNNKVVNIYVQTOYEELDPFRSRNVNKDIFICEHLNKLHLNTFVSEYL 203
QY 118 TIADIAVMSNLAGI--GQWESLRKSKYQNLVWFNSIDS-----EYKEALNEVVA 168
DB 204 TISDIFIFQYKYPHVSSTQYAKSKYKINRWHKLIISLIYEDAEWKNKILCLIC- 262
QY 169 FVGRGIGKSPAPSLKEKVHSD-----PSAPEVD-----LPGARV 206
DB 263 -----DNKKNPDRKDNLNKNNKSNNTDVKCKKVTTSYSGKLENAVIG 309
QY 207 KVCVAFAPESGYLIHGAKAALINKYFAERYQGLIYRFDPTNPSKESNEFEVULKD 266
DB 310 NVVTRFPPEPGYLIHGAKAAPLNYYAQMVEGMLRFDPTNVLELDIKEXSIIDL 369
QY 267 ETLGIKYAVTYSIDYFPELMEMASLIKQKAYIIDPTPEQMRKERMDGISRCRNNTV 326
DB 370 ENLGKYEKISYSSHPFLLEKCYCIDMIKMKAVADTGVEDMRQGEIESINRNSI 429
QY 327 EENSLAKEMVNGTERGQCCVRGLDMQDPNKSIRDPIYVRCNTDPHHRVGSXYKYP 385
DB 430 EKMLEFEMERKGTIEGQKNCIRAKINQSKCRDPVMTKCIYDVPHHKHOFKICYP 489
QY 386 TYDFACPFVDALEGVTHARSSHYHRNAQYRIIQQMGIRKVEIYFSLNMYTLLSK 445
DB 490 TYDFACPIIDIEGTHARINEYSRDEIQNWPFISTLNKRYIYERSRLAFVTVWSK 549
QY 446 RKLIMFVONKVEWTDPRPTVOGIVRGLKVEALIOFLQOGASKNLNMEMDGLMTI 505
DB 550 RKLKWFVENNVVDSVDRPPTIKGILRGLTKEALFQFILEQGPSKAGNLMQMDKWSI 609
QY 506 NKIIIDPVCARHTAVLKQRIYFTLTNGEPEEPVRLIPRHKKFEGAGKATTPANRIWD 565
DB 610 NKIIIDPPIPYAAVANKSSILLITLTDVIOQERDLHKNKSLGTCNNYNNKYIIE 669
QY 566 YADAAAIKGEVTLMDGNAIYKEIKVESGVITELVGBLHESGVKTKTKITWLADI- 624
DB 670 LEDAQTLLENSEITLILGNIIIKIEKENGKIQINALSNFHDGFKTTKKIHLPLP 729
QY 625 BELVPLSVEFDYLLSKKLEED--EDFLDNLPCTRRRIIPALGDANNRNIKGGIIQL 682
DB 720 QOLITCTLYEYDHLITVDKFNENDKDDWTNFINFNSKIEFTLVVAEPPSSLSKVDSPQFE 789
QY 683 RKGYYRCAPPIRSSKPVVLPALPDGRQQA 711
DB 790 RRGYFILDK-IDPHHLHLIKIPDGSK 816

RESULT 11

Q8SSE4 PRELIMINARY; PRT; 642 AA.
AC Q8SSE4;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE Glutamyl tRNA synthetase.
GN EC02.12.10.
OS Eucephalitozoon cuniculi.
OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Eucephalitozoon.
ON NCBI_TaxId=6035;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA Genoscope;
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA MEDLINE=21576510; PubMed=11719806;
RX Kacina M.D., Duprat S., Cornilloc E., Metenier G., Thomarat P.,
RA Kacina M.D., Barbe V., Peyretallade E., Broctier P., Wincker P.,
RA Delbac F., El Aloui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivas C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Eucephalitozoon cuniculi";
RL Nature 414:450-453(2001).
DR EMBL; AL590442; CAD25150.1; -
DR InterPro; IPR004526; GLX arch.
DR InterPro; IPR000924; Glu tRNA-synt_1c.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00749; tRNA-synt_1c; 1.
DR Pfam; PF03950; tRNA-synt_1c; 1.
DR TIGRfam; TIGR00463; glx arch; 1.
DR PROSITE; PS00178; AA TRNA LIGASE I; 1.
SQ SEQUENCE 642 AA; 73963 MW; B2BC6D4B2C329BB2 CRC64;

Query Match 30.5%; Score 1139; DB 5; Length 642;
Best Local Similarity 44.3%; Pred. No. 3.9e-73;
Matches 235; Conservative 92; Mismatches 184; Indels 20; Gaps 8;

QY 177 KSPAPSLKE-KVHSDKDPAPPEVDLPKAKVKVCVRPAPEPGYLIHGHAKALLNKYPA 235
DB 120 KANRFLKEFNAGSKKEGNELEIGPSE--NVVTRPPEPGRHGHIGHARALNLNYFA 176
QY 236 ERYQRLIVRPDPTNPSKESNEFENLKDLETLGIKYDAVYTSDFPKLMEASLIK 295
DB 177 SKNGRLLVFPDPTNPSKEERFERGILSDLSLGINETLSHTSDYFDKIIDLVPLIG 236
QY 296 QGKAYIDTPEKQMKERMDGIESRCRNVTENSLMKEMVNGTERGQCCVVRKGLDMQ 355
DB 237 ESKAYADNTPQEVMDERGRGVESRCRSDVESKRIFEMARGNASGY--CLRAKIDMS 294
QY 356 DPNKSLRDPVYRCNTDPPHRYGSKYKYPYDPAFCPPVDALGEGYTHALRSEYHDRNAQ 415
DB 295 SSNRKMRDPVLFVRESPIHRTGDKYKYPYDPAFCPIVDSLEGLTSLRANERYDRNQ 354
QY 416 YRILQDMGLR-RVEIYEFSSRLNMYTLLSKRKLMPYONKKEVEDWTPRPFTVGIVAR 474
DB 355 YVWFIDNLRNRPRIHPSRLNPFNTVLSKSKKLKYVDNGVSGMDPRLATINGIRL 414
QY 475 GLKVEALLQFIQOGASKNLIMEDKLTINKKIIIDPYCAHTVAVLKQRIYFTLNGP 534
DB 415 GNMNEALREYILMOGVSOIKTCISMDKWAIRKKIDIPVAFVCOQDAVEASIDWTS 474
QY 535 EEPFRILPRHKKFEGAGKATTFANRIWLDVADAANKGEEVTLMDGNIVELIKVE 594
DB 475 E--YMDVPKHKKNDLGTKEVYSSQILLQEDGRVADNDEEFTLMNGNIVSKTVE 532
QY 595 SGVIELVGLHLESGSVTKTLKLTWLDIEBELVPLSLVEFDYLLSKKLEDEDFLDNL 654
DB 533 NCTVTKMEVSLNPDODFKLTNKMGSVSK-RGSYVELLAELVGNLMD--EDTEDLARF 588
QY 655 NPCTREIIPALGDANMRNIKGEIIOLEKGYRCD-----PFISSK 698
DB 589 NRNSVKEWYAESAIINVRGEVLIQFERNGFYCDGFLVNLPLFTQKR 639

QY 199 DLPKAKVKVCVRPAPEPGYLIHGHAKALLNKYFARVYQRLIVRPDPTNPSKESNEF 258
DB 19 DLASGKHTVTRPPEPBNGLIGHAKSLINFGIADYQGCNLRPDPTNPKEDIEY 78
QY 259 VENLKDIETGKIKDA-VTTSDFPKLMEASLIKQKAYIDTPEKQMKER----- 313
DB 79 VDSINDVEMGFHSGNVRSSDYFDQHLAYIELINKGLAYDELTPQORERGTLT 138
QY 314 MDGIESRCRNVTENSLMKEM-VNGTERGQCCVVRKGLDMQDNKSLRDPVYRCNTD 372
DB 139 QPKNSPYRDRSVEENLALFEKMTGGEFG-KACLRAKIDMASPFIVMRDPVLYRIKFA 197
QY 373 PPHRGSKYYKYPYDPAFCPPVDALGEGYTHALRSEYHDRNAQYRIIQLDMGLR-RVELY 431
DB 198 EHHQGNKWCYIPWDFHICISDALGEGYTHSLCTLEFODNRLYDVLNITIPVHPRQY 257
QY 432 EFSRLNMYTLLSKRKLMPYONKKEVEDWTPRPFTVGIVARGLKVEALLQFIQOGAS 491
DB 258 EFSRLNMYTLLSKRKLMPYONKKEVEDWTPRPFTVGIVARGLKVEALLQFIQOGAS 491
QY 492 KNLNLMEDKLTINKKIIIDPYCAHTVAVLKQRIYFTLNGPPEPFRILPRHKKFEGA 551
DB 318 KQDNTEIASLESICREDLNEMAPRAVIVPVLKIVENYQEGE--NVTMHNHKNDEM 375
QY 552 GKATTFANRIWLDVAD-AAALNK-----GEVTLMDGNA-IVKIKVSSVITELV 602
DB 376 GSRQVPFGEIWIADREBANQYKRLVIGKEVRL--NAVYIKAEVEKDA----- 427
QY 603 GELHSGSVK-----TKLK--ITWLDIEBELVPLSLVEFDYLLSKK 643
DB 428 -----EGVITTFCTYDADTSLKPADRGRKYGKGVHWSAALH-PVELIRLYDLRFSVFN 481
QY 644 LEEDEDFLDNINPCTREIIPALGDANMRNIKGEIIOLEKGYRCDAPFISSKPV 700
DB 482 PGAADDFLSVINPELVIKQGAPEPSLKDVAAGKAFQEREGYFCLDSRHSTAEKPV 538

Query Match 20.1%; Score 750; DB 16; Length 554;
Best Local Similarity 35.0%; Pred. No. 2.7e-45;
Matches 188; Conservative 92; Mismatches 205; Indels 52; Gaps 14;

QY 199 DLPKAKVKVCVRPAPEPGYLIHGHAKALLNKYFARVYQRLIVRPDPTNPSKESNEF 258
DB 19 DLASGKHTVTRPPEPBNGLIGHAKSLINFGIADYQGCNLRPDPTNPKEDIEY 78
QY 259 VENLKDIETGKIKDA-VTTSDFPKLMEASLIKQKAYIDTPEKQMKER----- 313
DB 79 VDSINDVEMGFHSGNVRSSDYFDQHLAYIELINKGLAYDELTPQORERGTLT 138
QY 314 MDGIESRCRNVTENSLMKEM-VNGTERGQCCVVRKGLDMQDNKSLRDPVYRCNTD 372
DB 139 QPKNSPYRDRSVEENLALFEKMTGGEFG-KACLRAKIDMASPFIVMRDPVLYRIKFA 197
QY 373 PPHRGSKYYKYPYDPAFCPPVDALGEGYTHALRSEYHDRNAQYRIIQLDMGLR-RVELY 431
DB 198 EHHQGNKWCYIPWDFHICISDALGEGYTHSLCTLEFODNRLYDVLNITIPVHPRQY 257
QY 432 EFSRLNMYTLLSKRKLMPYONKKEVEDWTPRPFTVGIVARGLKVEALLQFIQOGAS 491
DB 258 EFSRLNMYTLLSKRKLMPYONKKEVEDWTPRPFTVGIVARGLKVEALLQFIQOGAS 491
QY 492 KNLNLMEDKLTINKKIIIDPYCAHTVAVLKQRIYFTLNGPPEPFRILPRHKKFEGA 551
DB 318 KQDNTEIASLESICREDLNEMAPRAVIVPVLKIVENYQEGE--NVTMHNHKNDEM 375
QY 552 GKATTFANRIWLDVAD-AAALNK-----GEVTLMDGNA-IVKIKVSSVITELV 602
DB 376 GSRQVPFGEIWIADREBANQYKRLVIGKEVRL--NAVYIKAEVEKDA----- 427
QY 603 GELHSGSVK-----TKLK--ITWLDIEBELVPLSLVEFDYLLSKK 643
DB 428 -----EGVITTFCTYDADTSLKPADRGRKYGKGVHWSAALH-PVELIRLYDLRFSVFN 481
QY 644 LEEDEDFLDNINPCTREIIPALGDANMRNIKGEIIOLEKGYRCDAPFISSKPV 700
DB 482 PGAADDFLSVINPELVIKQGAPEPSLKDVAAGKAFQEREGYFCLDSRHSTAEKPV 538

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Glutamyl-tRNA synthetase, putative (EC 6.1.1.18).
GN PF13_0170.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,
RA Berriman M., Pain A., Hall N., Ackin R., Chillingworth C., Doggett J.,
RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrett B.,
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844509; CAD52457.1; -
KW Aminocyl-tRNA synthetase; Ligase.
SQ SEQUENCE 918 AA; 108531 MW; 1CBABA4689C80A8P CRC64;

Query Match 18.4%; Score 685.5; DB 5; Length 918;
Best Local Similarity 27.9%; Pred. No. 2.5e-40;
Matches 199; Conservative 108; Mismatches 244; Indels 163; Gaps 19;

QY 121 DIAVMSNLAGIGQRES-----LRKSK-----KYQVLVWFNSIDSEYKEA 161
DB 191 DMTWKKIFFFGQKKRTEPLDPLFLFYLLKKNHFFEEETRYINLHMF----- 240
QY 162 LNEVAAVFGKRGIGKSPAPSLKEKVH-----DSKDPASAP-----EVDLPGAKV 205
DB 241 -----LKKKIYDISNDKEDNTPSACNTNFIQIIEEDLAKKH 279
QY 206 GKVCVRPAPSPSGYLHIGAKAALLNKYFAERYQRLIVREDTNPSSKESNEFVENLLKD 265
DB 280 THVITRFPBPBNYGLHGHAKSICLNFGLSNKYGRTLRPDTPVTEIRIYESIKED 339
QY 266 IETGIRK-DAVTYSDFPKLMEASLIRQKAYIDDTPEQKRKEMD-----GTSR 320
DB 340 VKMLGYDKHEHLYFASNYFEOLYEWAKLILKOGDAYVDOSIEETRKRNKGNLKPQGVDP 399
QY 321 CRNNTVEENLSLMEKMNVTGTERGMOCCVRGKLDMDPKSLDPVYVRCNTDPHHRVSK 380
DB 400 YRNTVEENLQLFEMKNGLYKEGKYLRAKINMKSMMNLRDPLLYIMTKIKHKTQK 459
QY 381 YKVYPTYPACPFVDALBEGVTHALSSSEYHDNRNAYRILQDMGLRVEIYEFSLNMY 440
DB 460 WVIYPMVYAHQSDSIKITHSICLTFETHRPLYEMFQEKHLFKTRQIEFALNTVY 519
QY 441 TLLSKRLIMFYQKKVEDWTPRPYQGIYRRLKYBALIQFLQOGASGNLNLMD 500
DB 520 MWSKRKLTLVNEKYVDMDPRMPTISGMRRGYSPEDAIKDFCNKVGIAKRENMIPID 579
QY 501 KLMTINKKIIDPVCARHTAVLKDQVIFLTNY-GPEEPFVR--ILPRHKEEGAKKATT 557
DB 580 LLEPCVRDMDKKAIRLFAILKPLKVIITNYDADYEHINLVASNPKNQOMQFRLK 639
QY 558 FANRIWLDYADAALINKG-----EEVTLMDQNALIVEIKV--- 593
DB 640 FEKEIFIDHDFQEIPODNFRLANRYRLRYACITCNBEIYKDDQKVI--ELRCYVD 697
QY 594 ---BSGVTELV-----GELHLGSVKTYTKLKT--WLADIEELVPLSTVEPYL 638
DB 698 PSSSGSLCTNOKKUNVNOGQGDGDKOGGDKNKVKATIHWC-AKNSLQAEFFMYKL 756
QY 639 ISKKKLEDED-----FLDN-----INPCR 660
DB 757 FTKPBPESNEDETVQKITSTAHLENNNTNNOGNGKDEYASDTLLQDEHLEDSKAGMRK 816
QY 661 EI-----PALGDANRNIRKEGIIQLEKGYRCDAPFRSSKPVVLAIP 706
DB 817 YINNSLIHKGIVENYSTRFKIGDPIQERVGFTTKXD--TINLPLFNPILTV 869

RESULT 14
Q9LQ07 ID Q9LQ07 PRELIMINARY; PRT; 791 AA.

AC Q9LQ07;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE ESTs AU097578(E0618).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC clone, p0510F03."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002486; BAB03361.1; -
DR HSSP; P00962; 1GTR.
DR Gramene; Q9LQ07; -
DR InterPro; IPR004514; GINS.
DR InterPro; IPR000924; Gln_tRNA-synt_1c.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00749; tRNA-synt_1c; 1.
DR Pfam; PF03950; tRNA-synt_1c; 1.
DR PRINTS; PR00987; TRNASYNTHGLU.
DR TIGRfam; TIGR00440; GINS; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_1; 1.
SQ SEQUENCE 791 AA; 89345 MW; F42BBE8179EBBAOC CRC64;

Query Match 18.0%; Score 673.5; DB 10; Length 791;
Best Local Similarity 27.2%; Pred. No. 1.5e-39;
Matches 211; Conservative 139; Mismatches 300; Indels 125; Gaps 20;

QY 2 EAALSPSSDPSPT-SITCAATVGLPLTINSLAAGSAPTLQFAGSGLHGVNPIIL-YI 59
DB 30 ENALVNSKYTANLAVIAIEAGISGCDKTVGNLTVAT--KYPTNALVH--RPVLIDYI 84
QY 60 ARGASIASLSKNDIEFGHVWEL-EVAPFTLSGSEFENACLPVDFGLASTFLVGHGT 118
DB 85 -----VSTKNPQDLALSLFNTGPDSDLTGKFEBAAC-----GAGVV 122
QY 119 IADIAVMSNL-----AGIQRESLSKSKYQVLVWFNSIDSEYKEA 161
DB 123 VSTIEISTVAVNLKHNMEALLBQRYHNINVGNLGQVAKRKPMDKATKQIEDKGLAEI 182
QY 162 L-----NEVAAVFGKRGIGKSPAPSLKE----- 185
DB 183 LGPRTDADNVKPVKKKKEKAKVBEKKAAVTTAAPSEELNPYSIFQPEENFKVHTEI 242
QY 186 -----KVHDSKDPASPEVDLPGAKVGRKVCVAPAPSPSGYLHIGAKAALLNKYFAR 237
DB 243 FVSDGNITWRANNSKEIEKHLKATG--GKWTFFPPBPBNYGLHGHAKMFIPLGLAKE 299
QY 238 YQRLIVAFDDTNPSSKESNEFVENLLKDIEITGKYVDVTSDFPKLMEASLIRQK 297
DB 300 RNGCYLRFDDTNPBAEKKEIYIDHIQELVHMGSEPYVYITTSDFQALYEHAYELLKKG 359
QY 298 KAYIDTPKEQMRKERMKGISRCRNNTVEENLSLMEKMNVTGTERGMOCCVRGKLDMDP 357
DB 360 LAAYDHQTAEBEIKYREKKNNSPWRDRIEESLKLFEEMRGLLAEGAATLRMKQDMND 419
QY 358 NKSLRDPVYTRCNTDPHHRVSKTKVPTYPACPFVDALBEGVTHALSSSEYHDNRNAY 417
DB 420 NKNMSDLAVYRIKFTPHPADKWCIPSYVYAHQMDSLNITHSLCTLEFDIRRPSY 479
QY 418 RILODMGLRVEIYEFSLNINVTLLSKRLWFGQNKVEDWTPRPYQGIYRRLKYBAL 477
DB 480 WLLVALGLYQPYVWEYSRLNLSNTVMSGRKLNRLVTEKMGVMDPRLILLAGLRKGV 539
QY 478 VEALIOFLQOGASGNLNLMDKLTINKKIIDPVCARHTAVLKDQVIFLTNNGPEEP 537
DB 540 STAINSFICIGITR--SLIRVRLLEVYHIRELNKTSRAVWVNLPLKV--INTLEDEK 595

QY 538 FV---RLPRKKKEGAGKATTFRANRIMLDYADAAAIKCEVTLMDMGAI----- 587
Db 596 VIDLDGKMPADPADASSYKVPSPRIYIEKTFRLKSDKYGLAPGKSALLRYAFP 655
QY 588 VKELVESGVITELVGEHLH-GSVKTTKLK--ITWLADIEBELVPLSLVE---PDLISK 641
Db 656 IKCEIVVYVGDNDPDIIEIRAEYDPSKTTKPKGVLMVWQAPAGVPLVAVRLFPLKFLS 715
QY 642 KLEBEDFLDNLNCTRR-----EIPALGDANNRNIKRGEIIQLERKGYRCD 690
Db 716 ENPALBEDWGLDNLNSKEVIGAYAVPSLATAVL-----GDKFQFERLGYRAVD 765

RESULT 15

Q8EG26 PRELIMINARY; PRT; 556 AA.
ID Q8EG26
AC Q8EG26;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Glutaminyl-tRNA synthetase.
GN GLNS OR SOI786.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Mehe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Knout H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.,
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015623; AAN54839.1; --.
DR TIGR; SOI786; --.
KW Aminoacyl-tRNA synthetase; Complete proteome.
SQ SEQUENCE 556 AA; 64103 MW; 3A7E689DACCACAF60 CRC64;

Query Match 17.9%; Score 669; DB 16; Length 556;

Best Local Similarity 34.7%; Pred. No. 1.8e-39;
Matches 182; Conservative 77; Mismatches 223; Indels 42; Gaps 13;

QY 199 DLPGAQKVCVRFAPSPSGYLIHGAALINKYFAERYGRLIVRPDDTNPSEKSNF 258
Db 21 DLKSGKHTNVQTRFPPENGVIHGHAKSICINFGIARDYQGLCNLRFPDNTPEKEDIDY 80
QY 259 VENLIKDIETLGIKYDA-VVYTSDFPKLMWAESLIQKAYI---DTPKQMKRER 313
Db 81 VNSIQADVRLWLGFGWDGEVRYSSNYFDQLHYAVELINKGLAYVCFLNADREYRGTLK 140
QY 314 MDGISRCRNNVVENLSIMKEMVNGTERGMQCCVRGLMDQDPKSLRDPYVRCNDP 373
Db 141 EPGKSPYRDPVVEENLRFGMRLGEEFGECALRAKIDMASPFMCMDPVIYIRIPAH 200
QY 374 HHRVSGKTKVPTYPFACPFVDALGVTALRSSEYHNRNAQYRILODM---GLRRVE 429
Db 201 HKQTDDKCIYPMYFTHCISDLAHHITSLCTLEFQDNRLYDWLNDLDDFOAPNRTR 260
QY 430 IYEFRLMNVYTLSSKRLKLVQNKRYEDWTDPRFPVQGIVRGLKVEALIOFILOQ 489
Db 261 QYEFRLNLVEYTLMSKRLKLVTRKLVGWDDPRMPTIAGLRRGYTPASIRFCQRIQ 320
QY 490 ASKNLNLMEWDKLTINKKIIDPVCARHTAVLKQQRVIFTLTNGPEEPPVRI-LPRKKF 548

Db 321 ITKQENMIAMLDACIRBELNEHAPRAMAVLRPLKV--IENYPEGQIETIQASHPSD 378
QY 549 EGAGKATTFRANRIMLDYAD-AAAIK-----GEEYTLMDMGNAIYKEIKVE----- 594
Db 379 ESMGTRELAFCREIFIDVADFREBANRQYKGLVNGKEVRL--NAVY--TKAERCDS 433
QY 595 SGVIT-----ELVGEHLGSGVKTTLKITWLADIEBELVPLSLVEFDYLISSKLE 646
Db 434 EGNITTYCSYADDTLGNPADG--RKVGVITHV-EATTAKPAQFRLYQRLFTDPNPA 490
QY 647 DEDFLDNLNCTRRREIPALGDANNRNIKRGEIIQLERKGYRCD 690
Db 491 AETVDEVLPNSLEVNGLVASLANAPAEKAYQFEEREGYFCAD 534

Search completed: January 25, 2004, 17:13:14
Job time : 198 secs